UNIVERSIDADE
BEIRA INTERIOR

Speech and voice assessment in Parkinson's disease

Versão Final após defesa

Daniela Borges Xavier

Dissertação para obtenção do Grau de Mestre em
Bioengenharia
(2º ciclo de estudos)

Orientador: Prof.a Doutora Virginie dos Santos Felizardo

Agosto de 2024.

Declaração de Integridade

Eu, Daniela Borges Xavier, que abaixo assino, estudante com o número de inscrição M12929 do Mestrado em *Bioengenharia* da Faculdade de *Engenharia*, declaro ter desenvolvido o presente trabalho e elaborado o presente texto em total consonância com o **Código de Integridades da Universidade da Beira Interior**.

Mais concretamente afirmo não ter incorrido em qualquer das variedades de Fraude Académica, e que aqui declaro conhecer, que em particular atendi à exigida referenciação de frases, extratos, imagens e outras formas de trabalho intelectual, e assumindo assim na íntegra as responsabilidades da autoria.

Universidade da Beira Interior, Covilhã 01 /08 /2024

Dedication

Dedico esta dissertação a todas as pessoas presentes nesta etapa da minha vida, mas em especial à minha familia.

Acknowledgements

Ao terminar esta dissertação, venho expressar os meus mais sinceros agradecimentos.

Em primeiro lugar, quero agradecer à Prof.a Doutora Virginie dos Santos Felizardo, Universidade da Beira Interior, pela disponibilidade e valioso apoio que forneceu ao longo deste último ano.

Também expresso o meu agradecimento por todos os meus amigos que estiveram ao meu lado durante todo este percurso de longos cinco anos e me demonstraram o seu grande apoio.

E, por fim, quero agradecer à minha família, por tudo o que fazem para que esta fase seja possível e por estarem sempre ao meu lado quando é mais preciso.

Resumo

A doença de Parkinson é uma doença neurodegenerativa que afecta a coordenação motora do doente. Caracteriza-se pela presença de sintomas como tremores, rigidez corporal, dificuldade em andar, fadiga e dificuldades na fala, podendo também ser observados sintomas mentais como depressão e/ou ansiedade. Os problemas de fala, ou disartria, relacionados com o enfraquecimento dos músculos responsáveis pela fala, estão presentes em cerca de 90% dos doentes diagnosticados com a doença.

O diagnóstico da doença, que se baseia nos sintomas do doente, é muitas vezes feito tardiamente, o que pode influenciar o controlo futuro dos sintomas da doença. Assim, propõe-se uma abordagem para o diagnóstico e prognóstico da doença de Parkinson utilizando técnicas de *machine learning* através da avaliação de características da voz e da fala do paciente, características essas encontradas na literatura, como a entropia, a dimensão fractal, a fonação, a prosódia e a glotal.

Com esta abordagem, será analisada uma gravação da voz do paciente e extraídas e selecionadas as características presentes na voz e, utilizando tecnicas de *machine learning*, será efetuada a sua classificação. No caso do diagnóstico, ocorrerá a distinção entre o facto de o paciente ter ou não a doença e, no caso do prognóstico, a identificação do nível da doença.

Palavras-chave

Voz e Fala, Doença de Parkinson, Diagnóstico, Prognóstico, *Machine Learning*

Resumo Alargado

0.1 Introdução

Nesta secção é apresentado um resumo alargado da dissertação intitulada "Avaliação da fala e da voz na doença de Parkinson" ("*Speech and voice assessment in Parkinson's disease*"). A mesma encontra-se dividida da seguinte maneira: em primeiro lugar uma introdução, onde são apresentados e descritos o tema, objetivos e respetiva organização desta dissertação; em seguida é realizado o Estado da Arte, onde é apresentada uma Revisão de Literatura Sistemática, englobando os estudos e pesquisas encontrados na área da avaliação da fala e da voz para a doença de Parkinson; em seguida é ilustrada a abordagem proposta seguida dos resultados obtidos e posterior discussão; por fim são exibidas as conclusões finais e também possíveis futuros trabalhos.

0.2 Objetivos e contexto

A doença de Parkinson é uma doença neurodegenerativa que afeta principalmente a coordenação motora do paciente. Esta é caracterizada pela presença de diversos sintomas, tais como o tremor, a rigidez corporal, a dificuldade de locomoção, a fadiga, a dificuldades na fala, presente em cerca de 90% dos doentes, entre outros, dado o enfraquecimento e possível perda do controlo do sistema muscular relacionados com a diminuição dos níveis de neurónios dopaminérgicos na Substância Negra.

Esta neurodegeneração não apresenta uma cura, mas a mesma pode ser controlada, tentando fornecer uma melhor qualidade de vida ao paciente, dependendo do tempo até à ocorrência da deteção. A deteção da doença de Parkinson é, na maioria das vezes, realizada tardiamente, dado que o diagnóstico é maioritariamente baseado na presença ou ausência de sintomas, algo que pode aumentar as chances da ocorrência de falsos diagnósticos, pois, muitas vezes, os sintomas não são percetíveis e/ou são confundidos com outra doença.

Tendo isto em ponderação, o objetivo desta dissertação é, através da análise da voz do paciente, realizar o diagnóstico e eventual prognóstico da doença de Parkinson utilizando técnicas de aprendizagem automática. Para que esta abordagem fosse elaborada foram consideradas questões tais como:

- É possível diagnosticar a doença de Parkinson apenas através da fala do doente?
- Através da avaliação da voz, é possível fazer um prognóstico fiável dos diferentes níveis da doença?

- Que características, ou grupo de características, têm maior impacto no diagnóstico e prognóstico da doença?

Através da análise destas questões foi possível a formulação e definição dos objetivos para esta dissertação, que se resumem nos seguintes quatro pontos:

- A pesquisa e análise do atual Estado da Arte sobre possíveis questões em aberto.
- A formulação de uma abordagem, tanto para o diagnóstico como para o prognóstico da doença de Parkinson através da avaliação da fala e da voz.
- A possível identificação de características que possam apresentar um maior impacto no diagnóstico e no prognóstico da doença.
- Realização de uma comparação do desempenho para diferentes grupos de características.

Através destes objetivos apresentados, foi possível a realização de contribuições como uma Revisão de Literatura Sistemática que engloba os estudos presentes na área, a formulação de uma abordagem capaz de efetuar, tanto o diagnóstico como o prognóstico, da doença de Parkinson através do estudo da voz e da fala, e ainda um estudo de comparação sobre a importância e a influência das características da voz em estudo.

0.3 Estado da Arte

Neste capítulo é realizado uma pequena introdução ao tema, à doença de Parkinson e a características do discurso parkinsoniano, sendo também referidas fases comumente utilizadas em processos que fazem recurso a técnicas de aprendizagem automática, que podem ser resumidas em 3 diferentes etapas:

- Pré-processamento: nesta fase os dados são processados de forma a serem utilizados nas seguintes etapas.
- Extração e seleção de características: os dados previamente processados são analisados e as características mais importantes são extraídas e selecionadas.
- Classificação: as características selecionadas são avaliadas pelo modelo de aprendizagem e a sua classificação é efetuada.

Após estas introduções iniciais, é ainda realizada uma Revisão de Literatura Sistemática sobre as abordagens realizadas para o diagnóstico e/ou prognóstico da doença de Parkinson através da avaliação de características da voz ou fala do paciente. Esta revisão engloba os últimos cinco anos de pesquisas, tendo sido iniciada com 221 artigos únicos e concluída com uma amostra de 87 artigos. A partir dos desafios encontrados na mesma, foi formulada a abordagem a aplicar.

0.4 Abordagem para o diagnóstico e prognóstico da doença de Parkinson

A abordagem proposta é efetuada com recurso a dados provenientes da base de dados de livre acesso MDVR-KCL. Esta é constituída por 42 gravações de pessoas saudáveis e por 31 gravações de pacientes com a doença de Parkinson, das quais 38, do grupo saudável, e 27, do grupo de Parkinson, foram incluídas no estudo final.

Destas gravações foram extraídas características do foro da fonação, da prosódia, da glotal, da dimensão fractal e também da entropia, sendo realizados, em seguida, testes para a perceção da importância de cada característica em estudo.

Após a seleção das características mais importantes foi efetuada uma fase de classificação através do uso de 10 classificadores, ou seja, BG-KNN, BG(F)-DT, DT, SVM-SVC, AB, BG(T)-DT, RF, GB, LR e GNB.

Por fim, foi ainda realizado um estudo comparativo entre os diferentes grupos de características da voz em estudo, de forma a se obter e perceber o peso que cada um representa neste estudo.

0.5 Resultados e Discussão

Neste capítulo são descritos e discutidos os resultados obtidos aquando da aplicação da abordagem descrita anteriormente, apresentando-se divididos em duas fases, uma fase de diagnóstico e uma fase de prognóstico, sendo realizadas as mesmas etapas para ambos.

Em primeiro lugar, a fase de extração das características da voz é descrita, sendo possível a extração inicial de 177, número que foi reduzido para 168, dado a ocorrência de pequenas falhas para algumas gravações.

Em seguida, foram efetuados testes de avaliação de distribuição de dados e de compreensão da relação entre as diversas características, dos quais foi possível a seleção das características consideradas como mais importantes, diminuindo assim o número de características a ser utilizadas para cada cenário e fase aquando da etapa de classificação.

Por fim, foi efetuada a etapa de classificação dos dados. Esta divide-se em duas partes, uma parte inicial em que são classificadas as características mais importantes selecionadas na etapa anterior, e uma segunda fase em que é realizado um pequeno estudo sobre a importância e influência que cada conjunto de características apresenta para a classificação. Para este estudo comparativo, a etapa de seleção das características mais importantes não é realizado, sendo utilizadas todas as características adquiridas aquando da fase de extração.

Para a primeira etapa de classificação, na fase de diagnóstico, obtiveram-se resultados

muito promissores, uma vez que todos os classificadores foram capazes de fazer a distinção entre uma gravação de um paciente com doença de Parkinson e de uma pessoa saudável. Para a fase de prognóstico, apesar de se ter obtido resultados promissores, é de ressaltar também a falta de capacidade que alguns classificadores apresentaram para a distinção entre os diferentes níveis da doença.

Para a segunda etapa de classificação, através da análise do estudo presente no Apêndice A.5, foi possível a perceção de que certas combinações de características funcionam melhor do que outras, sendo que para algumas, a influência que têm depende tanto da aplicação, seja diagnóstico ou prognóstico, como também do classificador utilizado, podendo estas serem uma mais-valia, ou não, para o processo de classificação.

0.6 Conclusão e Trabalhos Futuros

Esta dissertação propõe uma abordagem para a doença de Parkinson, diagnosticando e prognosticando a doença através da avaliação de características da voz e da fala utilizando técnicas de aprendizagem automática.

Este tipo de abordagem constitui uma mais-valia, quer em termos de tempo, quer em termos de uma possível redução de falsos diagnósticos, para além de proporcionar uma visão do nível de doença do paciente, algo que pode complementar a importante análise efetuada por um profissional de saúde.
Assim sendo, foram alcançados quatro objetivos:

- Análise da atual literatura sobre possíveis questões em aberto.
- A formulação de uma abordagem para o diagnóstico e prognóstico da doença de Parkinson através da avaliação da fala e da voz.
- A identificação e seleção de características que apresentam um maior impacto no diagnóstico e no prognóstico da doença.
- Concretização de um estudo de comparação de desempenho para diferentes grupos de características.

Através destes objetivos foi possível obter contribuições importantes.

Através da análise da Revisão de Literatura foi possível notar que a grande maioria das abordagens apresenta apenas uma proposta de diagnóstico ou deteção da doença, sendo reduzido o número de abordagens para o prognóstico. É também verificado que as abordagens apresentadas avaliam principalmente gravações de vogais sustentadas ou de pequenos fonemas, e menos gravações de fala contínua.

Através do segundo objetivo foi proposta uma abordagem para o diagnóstico e prognóstico da doença de Parkinson através da análise e avaliação da voz e da fala através da entropia,

da dimensão fractal, da fonação, da prosódia e características glotais, da qual foi possível a obtenção de resultados promissores, podendo-se concluir que a abordagem proposta é viável e capaz de efetuar o diagnóstico e o prognóstico da doença de Parkinson.

Com o cumprimento do terceiro objetivo, concluiu-se que para cada cenário e aplicação, seja de diagnóstico ou de prognóstico, existe uma variação nas características escolhidas como mais importantes pelos testes, tanto em número como em tipo.

Para o último objetivo, um estudo comparativo entre os diferentes tipos de características e possíveis combinações entre elas, de modo a determinar o seu peso e influência nos resultados obtidos, foi conduzido. Este estudo permitiu concluir que a utilização de certos grupos, como a fonação, a prosódia e a glotal, são uma mais-valia para o processo de classificação. Concluiu-se também que a influência do grupo que englobam as características da entropia e da dimensão fractal é algo variável e depende da aplicação, seja ela de diagnóstico ou prognóstico, e do classificador utilizado, podendo ser uma mais-valia para uns e uma desvantagem para outros.

Com a conclusão deste estudo de pesquisa e dos respetivos objetivos, algumas questões e possibilidades para futuras investigações foram elucidadas.

A primeira questão é relacionada com a reprodutibilidade desta abordagem para a análise de gravações numa língua diferente da língua inglesa. A segunda questão é sobre a possível influência que o género, masculino ou feminino, dos sujeitos em avaliação poderá ter tido nos resultados obtidos, apesar de não ter sido considerado como variável.

Em termos de possibilidades, temos como primeira a realização desta abordagem, principalmente para a fase de prognóstico, utilizando um maior número de pacientes para os quatro níveis da doença. Em seguida, a realização desta abordagem utilizando uma combinação das características das matrizes estáticas e dinâmicas das características de fonação, de prosódia e de glotal, e também a reprodução desta abordagem efetuando uma etapa de pré-processamento. E por fim, uma possibilidade de implementar esta abordagem, por exemplo, numa aplicação móvel ou num dispositivo *wearable*, a fim de complementar o procedimento clínico praticado.

Abstract

Parkinson's disease is a neurodegenerative disease that affects the patient's motor coordination. It is characterized by the presence of symptoms such as tremor, body rigidity, difficulty walking, fatigue, and speech difficulties, and mental symptoms such as depression and/or anxiety may also be observed. Speech problems, or dysarthria, related to the weakening of the muscles responsible for speech, are present in around 90% of patients diagnosed with the disease.

The diagnosis of the disease, which is based on the patient's symptoms, is often made late, something that may influence future control of the disease's symptoms. Therefore, an approach is proposed for the diagnosis and prognosis of Parkinson's disease using machine learning techniques through the assessment of features of the patient's voice and speech, features found in the literature, such as entropy, fractal dimension, phonation, prosody, and glottal.

With this approach, a recording of the patient's voice will be analyzed and the characteristics present in the voice will be extracted and selected and, using machine learning techniques, classified. In the case of diagnosis, a distinction will be made between whether or not the patient has the disease and, in the case of prognosis, the level of the disease will be identified.

Keywords

Voice and Speech, Parkinson's disease, Diagnosis, Prognosis, Machine Learning

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

AB	Adaptive Boosting
ANN	Artificial Neural Networks
CNN	Convolutional Neural Networks
DFA	Detrended fluctuation analysis
DT	Decision Tree
GNB	Gaussian Naive Bayes
HFD	Higuchi fractal dimension
HRF	Harmonic Richness Factor
KNN	K-Nearest Neighbors
LR	Logistic Regression
MFCC	Mel Frequency Cepstral Coefficient
NAQ	Normalized Amplitude Quotient
OQ	Opening Quotient
PD	Parkinson's disease
PE	Permutation Entropy
RF	Random Forest
STFT	Short-Time-Fourier Transform
SVM	Support Vetor Machine
SVDE	Singular Value Decomposition Entropy

Chapter 1

Introduction

This chapter presents an introduction to the dissertation entitled ”Speech and voice assessment in Parkinson’s disease”.
This dissertation will address the diagnosis and prognosis of Parkinson’s disease (PD) through the voice and speech of patients of PD using machine learning techniques. The context and motivation for carrying out this research, the problematic and the objectives, as well as the main contributions of the study and the organization of this dissertation, are illustrated in this chapter.

1.1 Context and Motivation

Parkinson’s disease is a chronic neurodegenerative disease that affects motor coordination. It is characterized by the loss of dopaminergic neurons in the substantia nigra [1]. This movement disorder have no cure, but can be controlled, reducing the effects of the disease, by medication and therapies [1].
In 2019, it was globally estimated that 8.5 million people suffer from PD [2].
It has a higher prevalence of diagnosis in the male sex and in people in older age groups (>50 years), but this does not exclude its existence in younger age groups [2], or in the female community. Although there is no higher prevalence in this sex, when diagnosed, they can have higher risk factors for symptoms such as dyskinesia and possible fluctuations in motor and nonmotor response, related to overdosing on the therapies given, as well as a higher likelihood of having urinary complaints and depressive symptoms [3].
In addition to the extremely disabling physical symptoms, which lead to the patient becoming dependent on others in the more advanced stages of the disease, the disease also affects emotional and psychological levels and can lead to an increase in anxiety and depressive symptoms in this percentage of the population [1].
In terms of the symptoms of the disease, which will be explained in the chapter 2.1.2 ,more than half of patients suffer from some kind of speech difficulty, better known as dysarthria, something that can be observed in the early stages of the disease, helping in its diagnosis and also in a future prognosis [1].

1.2 Problematic and objectives

Currently, there is no specific test for detecting the disease, so diagnosis is based on clinical methods based on the patient's physical symptoms [4], which are usually only noticed at an advanced stage of the disease, making it difficult to diagnose and control the disease at an early stage. It is also possible to at least help diagnose through magnetic resonance imaging (MRI), in which a specific biomarker is sought [4], something that could be considered invasive for the patient.
The use of machine learning techniques can be extremely useful in this biomedical field, since certain symptoms and changes are not perceptible to the human eye or ear, but certain algorithms, when trained, are able to detect these small changes and help make a faster and more accurate diagnosis, thus reducing the number of false positives and false negatives that occur during clinical detection, as well as making a prognosis of the disease, as exemplified in Section 2.2.
Considering the current difficulty in diagnosing and prognosticating this neurological disease, this dissertation focuses on developing an approach to the diagnosis and prognosis of Parkinson's disease. For this approach, some research questions in the field were considered, such as:

- It is possible to diagnose Parkinson's disease solely by assessing the patient's speech?
- Through speech assessment, is it possible to make a reliable prognosis of the different levels of Parkinson's disease?
- Which features, or group of features, have the greatest impact on the diagnosis and prognosis of the disease? (taking into account a reduction in the computational time required)

Based on these questions, it was possible to formulate some goals for this investigation.

- Research and analysis in the literature review on possible open questions in the field.
- The formulation of an approach to the diagnosis and prognosis of Parkinson's disease through the evaluation of speech and voice.
- Identification of features with the most impact on the diagnosis and prognosis of Parkinson's disease through voice analysis.
- Compare performance values for different groups of features, reducing computational time, without negatively influencing the final results

1.3 Main contributions

Considering the objectives presented for this dissertation, the following contributions were achieved:

- a systematic literature review covering the approaches presented in the last five years in the area of diagnosis and prognosis of Parkinson's disease through speech assessment.
- an approach to the diagnosis and the prognosis of Parkinson's disease through speech evaluation using machine learning techniques
- a comparison between different machine learning classifiers, taking into account the performance values achieved.
- a comparative study of different feature groups and their possible impact on the diagnosis and prognosis of the disease

1.4 Dissertation organization

This last section of this chapter describes the organization of this dissertation. The dissertation is divided into five chapters, briefly summarized as follows.

- Chapter 1: the topic to be discussed in this dissertation is presented and described, as well as the motivations for performing this dissertation, its organization and its main contributions.
- Chapter 2: the state-of-the-art is presented, and the research method for the Systematic Literature Review of the proposed theme is elucidated, as well as the discussion of the research results.
- Chapter 3: the methodology applied is explained.
- Chapter 4: the results obtained are presented and discussed.
- Chapter 5: final conclusions and future work are outlined.

Chapter 2

State-of-the-Art

This chapter describes and discusses the state-of-the-art of Parkinson's disease diagnosis and prognosis through assessment of voice and speech, providing a background on the topic and a systematic review of the literature covering approaches in this field.

2.1 Background Knowledge

2.1.1 Human brain

To better comprehend PD, a neurodegenerative disease, it is important to understand their pathological characteristics and how the human brain is affected.
The brain is a fascinating and extremely complex machine that is responsible for controlling all the actions and functions of the human body. This organ can be split into three major areas: the forebrain, the midbrain, and the hindbrain [5].
The hindbrain encompasses the cerebellum, the upper part of the spinal cord and the brain stem, on the other hand, the forebrain include the cerebrum and sections such as the hypothalamus, the thalamus and the hippocampus. Midbrain is the smallest and is responsible for some voluntary movements and the eye movements [5], for example.
The control of the motor system is covered by the hindbrain and midbrain, more specifically, areas such as the cerebellum and the basal ganglia [5].
As mentioned above, PD is a motor disease that affects the nervous systems and consequently the patient's movements. In the case of Parkinson's disease, the basal ganglia are the most affected.
As referred to in Chapter 1, in this pathology there is a loss of dopaminergic neurons in the substantia nigra, an important structure of the basal ganglia, located in the midbrain [6].
Substantia nigra is a nucleus with an important role, mainly in the formation and supply of neurotransmitters to different areas of the brain. It is divided into two parts, the pars compacta and the pars reticulata. When one of these parts degenerates, specifically the pars compacta, there is a failure in neurotransmitter supplementation, in this case dopaminergic neurons die, which leads to a reduction in the supply of dopamine neurotransmitters [6]. These neurotransmitters are distributed through the action of certain proteins, such as the alpha-synuclein protein. Since neurons die and there is a subsequent decrease in the levels of dopamine being distributed, this will lead to the accumulation of this proteins. This accumulation results in Lewis bodies and Lewy neurites, which are

nothing more than an abnormal aggregation of α-synuclein protein [3], which leads, for example, to failures in the transmission of information by the motor system.
As to why this type of malfunction occurs in this part of the substantia nigra, there is currently no certainty. It can be attributed to genetic factors or be caused by external factors, such as exposure to environmental pollution or certain toxins, such as the neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), a heroin analogue [3].

2.1.2 Parkinson's disease

The first clinical records of Parkinson's disease date back to 1817 by James Parkinson, describing it as a neurological syndrome, although there are indian and chinese writings from 1000 BC that suggest the discreteness of PD [7]. At the time, the first symptoms described were the tremor and rigidity that the patients presented, and only later, in 1872, by Jean-Martin Charcot, that the existence of bradykinesia was pointed out, author who first suggested the use of the term "Parkinson's disease" to refer to the pathology [7].
As mentioned in Chapter 1, Parkinson's disease is a neurodegenerative disease that affects the motor coordination of patients suffering from it. Affecting the patient physically, psychologically, and emotionally.
This disease is characterized by the existence of certain symptoms that can be of two types, motor and non-motor symptoms. Motor symptoms include slow down of spontaneous and automatic movement or bradykinesia, tremor, involuntary movements, rigidity, loss of balance, painful muscle contractions, and consequently difficulty in locomotion. Non-motor symptoms include cognitive impairment, mental health disorders, dementia, sleep disorders, pain and sensory disturbances [2]. Patients can also discuss other problems related to these, such as hyposmia [3], urinary problems or constipation, skin problems, orthostatic hypotension, fatigue and loss of energy, sexual dysfunction, psychotic symptoms, cause by the medication, difficulty with swallowing and chewing, and speech changes [1].

2.1.2.1 Parkinsonian speech

As previously mentioned, changes in speech or dysarthria, are a common symptom in people suffering from PD, occurring in around 90% of cases [8]. Changes are noticeable in terms of tone and spontaneity of speech, with occurring a speech in monotone or too softly, an hesitation when starting or speaking too quickly [1], difficulty in word finding, impaired speech, prolonged speech time or pronunciation ambiguities [8].
Dysarthria is due to injuries to the motor neurological system that lead to weakness or paralysis of the muscles responsible for speech and breathing. This weakening of the muscles influences the phonatory and articulatory system, leading to a change in prosody and an increase in fundamental frequency, due to changes in vocal intensity and the vibration of the vocal cords, as well articulatory inaccuracies and an increase in the number of

pauses [8].
Changes such as those mentioned are taken into account when diagnosing and prognosing Parkinson's disease, and are used as features to identify the pathology when detected by machine and deep learning techniques, as can be seen in the following sections.

2.1.3 Parkinson's disease assessment

As already mentioned, certain characteristics of this neurodegenerative disease are not perceptible only by human hearing or sight, thus causing gaps and delays in diagnosis, something that will influence future treatment and attempts to reduce the symptoms of the disease.
We live in an age of technology and innovation and, without a shadow of a doubt, we need to take advantage of this. If we take a look at what has already been done in the area of diagnosis and prognosis of Parkinson's disease, we can see that the use of machine learning techniques has been increasing in recent years and with excellent success rates, which helps to reduce the number of incorrect diagnoses.
Having said this, the diagnosis and prognosis of PD using machine learning techniques, although it can be carried out using different patient data, i.e. obtained through writing or drawing, the level of tremor, the patient's gait, speech, among others, all the methods have something in common, i.e. the data used needs to go through stages of processing, selection and classification, in order to obtain the best features and the best possible performance, taking into account the cost and computational time.
These different stages can be differentiated into 3 phases:

- **Preprocessing**: in this phase the raw data is processed and adapted to a machine learning model [9];
- **Feature extraction and feature selection**: the data previously processed is analyzed and the most important features are extracted and selected;
- **Classification**: the selected features are evaluated by the learning model and their classification is carried out.

2.1.3.1 Preprocessing

To use a machine learning technique and obtain the best possible results, it is necessary to work and adapt the data that will be supplied to the algorithm, so that it can obtain the necessary information without interference that could influence the final outcome. Therefore, in order to facilitate data evaluation, an initial stage known as preprocessing is necessary.
In this phase, the data is evaluated and worked on [9], and interference such as noise or unnecessary pauses, among others, are reduced or even removed if possible. It is at this

stage that changes are made to the sampling frequency or analysis time, such as dividing the signal into smaller fragments, in order to make the analysis easier or more objective. To make this possible, techniques such as normalization, segmentation, noise reduction, among others, are commonly used [10].

2.1.3.2 Feature Extraction and Feature Selection

After preprocessing, it is necessary to extract the information contained in the data signal, for which a second process called feature extraction is performed. In this process, the input data are analyzed and features are acquired, i.e. extracted, and then the most important features are selected, i.e. feature selection, reducing the number of variables without losing valuable data and possibly improving the overall performance of the chosen model [9]. Different techniques can be applied depending on the desired features of the signal, including the use of certain algorithms or software built for this purpose.
In terms of feature extraction, can be grouped into two methods [11]:

- **Supervised Methods**: In supervised methods, data are evaluated according to certain labels and classes, depending on the characteristics of the information acquired. As an example, we have the use of techniques such as Supervised Principal Component (SPC) or Linear Discriminant Analysis (LDA).
- **Unsupervised Methods**: In this method, the extraction does not take into account possible tabulations of the input data, but rather possible variations and the way in which the information is distributed and directed. As an example, we have techniques such as Principal Component Analysis (PCA) and the use of Autoencoders, a type of neural network.

The feature selection process can be divided into three methods:

- **Filter methods**: In this type of feature selection, as the name implies, filtering methods are applied to choose the features that will be provided for the future classification phase. These filtering methods are mainly based on the relevance of the data to the target and also on possible similarities between the information. Criteria such as Pearson's Correlation Coefficient (PCC), X^2 Statistics and Minimum-Redundancy-Maximum-Relevance (mRMR) are examples of this method. [11]
- **Wrapper methods**: In the case of wrapper methods, the best features are selected by a learning model. In this case, the method can be grouped into two types, depending on the algorithm: sequential method or metaheuristic method. In the sequential method, the selection is made according to a given sequence, with each feature individually tested to see whether it really adds value to the final performance. In the metaheuristic method, metaheuristic algorithms such as Modified Grey Wolf Opti-

mization (mGWO), Particle Swarn Optimization (PSO) or Genetic Algorithms (GA) are applied to select the features that improve performance. [11]

- **Embedded methods**: In the case of embedded methods, the features are selected when the classifier algorithm is built. In other words, this selection is made by algorithms such as Support Vector Machine or Random Forest, and depends on the choices made by the classifiers themselves. The selection made by one algorithm may not work for another type of algorithm. [12]

The type of audio features to be extracted and selected can be grouped depending on their domain. Taking into account the classifications presented in [13] and [14], the following domains can be highlighted:

- **Time/Frequency domain features**: Examples include zero-crossing rate features, Short-Time-Fourier Transform, tonality features, and amplitude-based features.
- **Wavelet/Cepstral domain features**: Examples are Mel Frequency Cepstral Coefficients (MFCCs) and Perceptual Linear Prediction Coefficients (PLP).

2.1.3.3 Classification

In the disease assessment process, the classification phase is important, as it is the culmination of the two previous phases. At this stage, it is essential to choose a prediction model that fits the classification objective.
Machine learning prediction models, i.e., models capable of learning and making classifications and predictions by evaluating patterns and values observed in input data [15], can be divided into three groups, according to the type of learning. Namely [15],

- **Supervised Learning**: Supervised Learning techniques are the most commonly used. In this type of technique, the input data is categorized and the desired output is known. Taking these variables into account, the algorithm establishes the necessary relationships between the characteristics provided and the desired output, learning the different patterns that exist by evaluating the training data and thus testing and improving its performance. This can then be used to discriminate between unknown data. This type of learning can also be divided into classification models and regression models. The former presents an output with a binary class, and the latter presents an output with a continuous value.
- **Unsupervised Learning**: In this type of learning, unlike the previous one, the input data provided do not need to be categorized, and the intended output may or may not be known. The algorithm will cluster the information provided according to certain patterns and similarities observed, providing an output based on these relationships established between the initial information.

- **Reinforcement Learning**: Reinforcement learning, in a nutshell, operates on the basis of trial and error techniques, as well as reward and punishment techniques. In other words, using the data provided as input, the algorithm will evaluate and classify the information and present an output. This output is evaluated to see if it was the right one or not. If it was the right one, the algorithm is ”rewarded” and the classification is considered a good prediction, but if the output is not the right one, the algorithm is ”punished”, and this attempt is considered an error. Through these ”rewards and punishments”, the algorithm adapts and tries to correct possible errors, improving the output presented.

In terms of the use of classification and prediction algorithms in the field of diagnosis and prognosis of neurological diseases, in this case Parkinson's disease, some machine learning algorithms stand out. These include algorithms such as Support Vetor Machine (SVM), Random Forest (RF), Decision Trees (DT), K-Nearest Neighbors (KNN), and Neural Networks (NN) [16] [17].
Support Vetor Machine is a supervised learning model that uses kernel-based functions performing binary classifications. As a supervised learning algorithm, it analyzes tabulated data and subsequently classifies it [17].
Decision Trees resembles the structure of a tree, as the name suggests. It is a decision algorithm that groups information into nodes or blocks based on certain attributes and similarities [17].
Random Forest is an algorithm presented by Breiman [18], it derives from DT. In this model, several decision trees are used as decision methods, where each variable is chosen randomly and analyzed, presenting the final output that more decision trees obtained [19].
In the field of Neural Networks use, Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) stand out.
Artificial Neural Networks are learning models that, as the name implies, try to imitate the functioning and connections of the human neuronal system. They are a category of algorithm that have become very popular in recent years. In their simplest form, they consist of three layers: an input layer, a hidden layer and an output layer. Each layer has one or more nodes, which simulate neurons, as well as weights. In this type of network, information only flows in one direction, passing from neuron to neuron from the input layer to the output layer [15].
Convolutional Neural Networks is an algorithm that, like Artificial Neural Networks, are designed to mimic a neural network. Unlike the previous one, this neural network is made up of multiple hidden layers, each of which plays a role in the learning process. Although it has better classification capabilities, a greater amount of data is required for it to operate correctly. This type of neural network also has a wider range of applications than simple ANNs [15].
For the classification process to be more accurate, in addition to choosing the right classifier, it is common to split the data. It can be divided into three groups[20], namely a

training group, usually with a higher percentage of data, which will be used to train the algorithm in terms of patterns and characteristics of the data, a second group called the validation group, which serves to improve the algorithm's classification, and a final group called the test group, which, as the name implies, serves to test the algorithm's level of learning in relation to the training data. In this last group the performance evaluation tests are carried out.

2.1.3.4 Performance metrics

In order to be able to check the success rate of the previously chosen classifier, taking into account the data used for training and its classification for test data, i.e. data unknown to the algorithm, there are certain ratio relationships that can be calculated. Using these ratios, it is possible to compare the results with other types of classifier and understand which is the most suitable, thus determining the best solution for the problem in question [21].
In summary, performance evaluation methods can be described as the ratio between the number of false positives (FP), true positives (TP), false negatives (FN) and true negatives (TN) obtained by the classifier.
Where,

- True Positives are the number of cases correctly classified as PD;
- False Positives are the number of cases incorrectly classified as PD;
- True Negatives are the number of cases correctly classified as healthy patients;
- False Negatives are the number of cases incorrectly classified as healthy patients.

Among the performance evaluation techniques, the following can be highlighted:

- **Accuracy**, presents the relationship between the correct predictions over the total number of cases [21]. Being calculated using the formula:

$$ACC = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (2.1)$$

- **Precision**, gives the ratio between the number of correctly classified cases and the total number of positive cases [21]. Being calculated using the formula:

$$Prec = \frac{TP}{TP + FP} \quad (2.2)$$

- **Sensitivity/Recall**, shows the percentage of cases that were classified correctly by

the classifier [9]. Being calculated using the formula:

$$Sen = \frac{TP}{TP + FN} \quad (2.3)$$

- **F-measure/F1-Score**, shows the relationship between the values obtained for precision and sensitivity [21]. Being calculated using the formula:

$$F1 - score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \quad (2.4)$$

- **Specificity**, shows the proportion of cases that were incorrectly classified [21]. Being calculated using the formula:

$$Spec = \frac{TN}{TN + FP} \quad (2.5)$$

Techniques such as mean absolute error (MAE) [22], area under the curve (AUC), and confusion matrix [21] can also be used to evaluate classifier performance in the assessment of Parkinson's disease.

2.2 Systematic Literature Review

This section present the Systematic Literature Review of this dissertation, as well as the research techniques used for its detailed realisation. First, the method for selecting the articles is presented, followed by a discussion of the systematic literature review presented.

2.2.1 Research Strategy

For this multidisciplinary research, was used SienceDirect, IEEE Xplore and ACM Library databases, from which was obtained a good range of articles and studies carried out in the proposed research area.
The following research is focused on titles, abstracts, and keywords, covering an interval of ten years (2014-2023), interval that was later reduced to five years (2019-2023), given the large number of articles included in the final selection.
The keywords used for the Systematic Literature review research were the followings:

- (parkinson disease) AND (speech OR voice) AND (assessment OR processing);

2.2.2 Article Selection

During the development of this systematic literature review, duplicate articles and studies in foreign languages were eliminated, articles that did not include a voice or speech

evaluation or that did not use machine learning tools for the diagnosis or prognosis of Parkinson's disease were also excluded.
The number of participants with Parkinson's disease included in the study was taken as a criterion, as well as the methods used to treat voice and speech, such as preprocessing, feature extraction, and classification techniques, as well as the final aim of the research, i.e. whether it was for diagnosis/detection or prognosis of the disease. After this selection, the resulting scheme is illustrated in the figure 1.

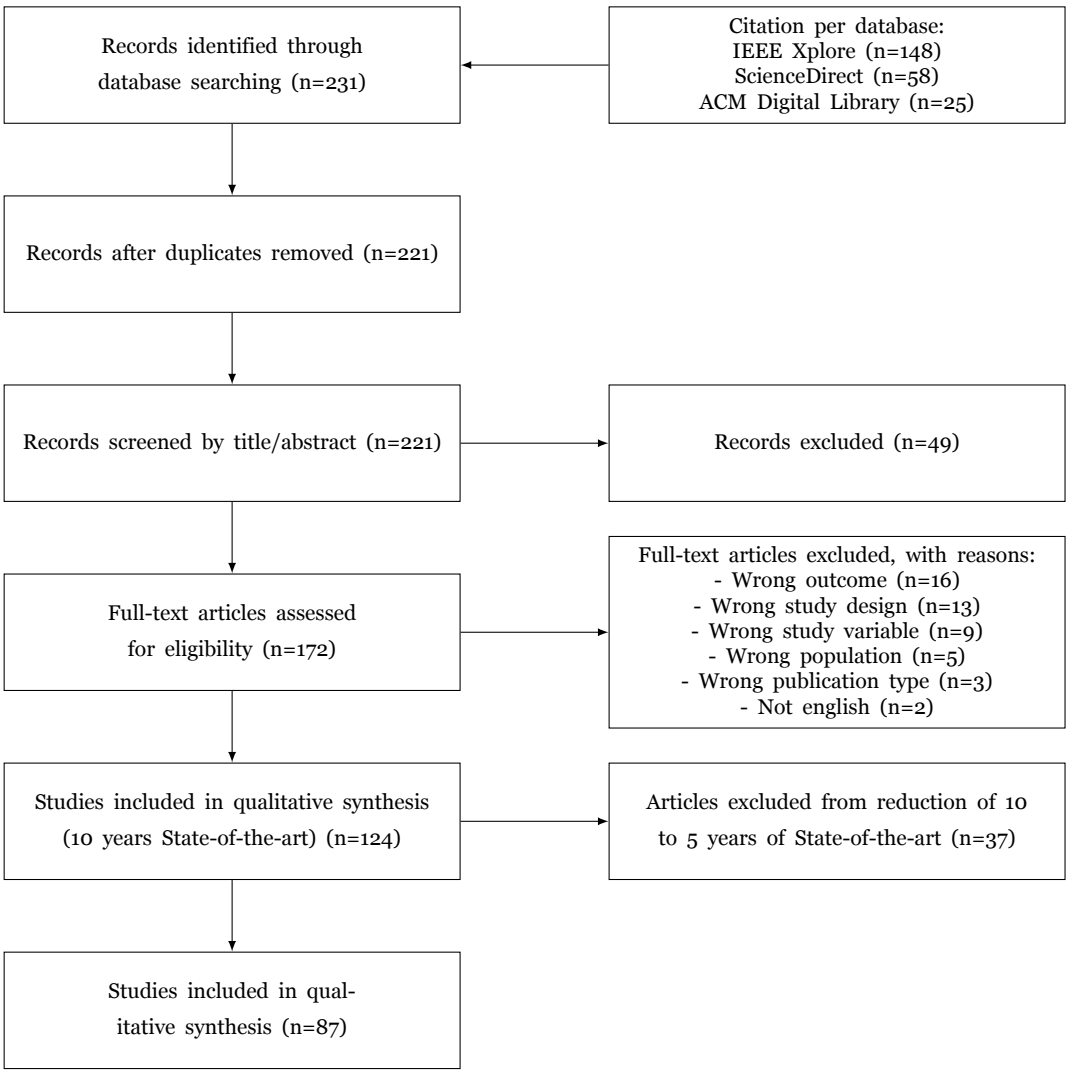

Figure 1: Article selection scheme

From figure 1, it is shown that the total number of articles collected from the chosen databases was 221. Taking into account the title and abstract of the articles, that number was changed to 172. After this, an evaluation of the articles in question was carried out, taking into account factors such as the number of patients and the variable under study, studies with fewer than 10 subjects with PD were excluded and articles using more than one variable were admitted, as long as one of them was the voice or speech of patients

with Parkinson's disease, the techniques used for the treatment and classification of the audio signal were also taken into account, as well as the final objective of the study. As a result, 124 articles were obtained, corresponding to 10 years of state-of-the-art, a high number, which was later reduced to 87 by reducing the number of years of state-of-the-art from 10 to 5. The records excluded are described as follows:

1. sixteen articles did not diagnose or prognosticate the disease;
2. thirteen studies did not mention the techniques used to process or classify the data;
3. nine studies did not used voice or speech signals;
4. five studies have less than ten PD patients or do not use PD subjects;
5. three articles are reviews or book chapters;
6. two studies are not in English;

The final selection comprises 87 articles, which are described in Table 1.

Table 1: Related works.

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[23]	2019	2 databases (Train+Test)	20 PD/20 HC + 28 PD	Praat software; Normalization	χ2 statistical model (noise elimination); SSFH algorithm	NN	LOSO CV	**ACC**:97.5%-100%; **Sen**:100%; **MCC**:0.951; **Spec**:95%;	Diagnosis/ detection
[24]	2019	Parkinson's disease dataset from UCI machine-learning database	188 PD	SMOTE (Synthetic Minority Oversampling Technique) (balanced class distribution)	753 features	SMOTE + Random florest	50/50 (training/test) 10-fold CV	**ACC**:94.89%; **F1-score**:0.949; **Kappa value**:0.894; **Prec**:95.1%; **AUC**:0.991; **Sen**:94.9%	Diagnosis/ detection
[16]	2019	UCI	188 PD/64 HC	Min-max normalization	Mel-Frequency Cepstral Coefficients (MFCCs); Wavelet transform (WT); Tunable Q-factor wavelet transform (TQWT); Glottis Quotient (GQ); Glottal to Noise Excitation (GNE); Vocal Fold Excitation Ratio (VFER); Empirical Mode Decomposition (EMD)	CNN + SVM	LOPO CV	**ACC**:86.9%; **F1-score**:0.917; **MCC**:0.632 (Model-level combination)	Diagnosis/ detection
[25]	2019	Department of Neurology, Istanbul University	188 PD/64 HC	Standardization; Min redundancy-max relevance (mRMR) (feature selection (FS))	Baseline Features; Time frequency features; MFCCs; Vocal fold features; WT; TQWT	SVM (linear+RBF); MLP; NB; LR; RF; k-NN	LOSO CV	**ACC**:86%; **F1-score:**0.84; **MCC**:0.59 (SVM-RBF classifier)	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[26]	2019	Neurovoz	52 PD/56 HC	Normalized, framed and parameterized RASTA-PLP	Perceptual Linear Prediction (PLP); MFCCs	Siamese LSTM-based NN		**EER**:2.9% -> sentence 1 all 5 sentences -> 1.9%	Diagnosis/ detection
[27]	2019	VGFR Dataset (gait); Voice Impairment Dataset - UCI (voice)	70 PD/55 HC (train) + 23 PD/18 HC (test) (gait) + 100 PD/40 HC (train) + 38 PD/17 HC (test) (voice)	VGFR Spectrogram Detector (gait); Normalization (standart) + weight measure (training data) (voice)	VGFR Spectrogram Detector (gait); Variations of Fundamental Frequency + amplitude - Voice Impairment Classifier (voice)	CNN (gait); ANN (voice)	Validation of VGFR Spect. Det. And Voice Imp. Class.; Comparison with XG Boost, SVM, MLP	**ACC**:88.17% (VGFR); 89.15% (Voice IC)	Diagnosis/ detection
[28]	2019	Collected Dataset (speech, handwriting, gait) + 3 speech dataset (PC-GITA)	40 PD/40 HC (speech, handwriting,gait) + spanish: 50 PD/50 HC + german: 88 PD/88 HC + czech: 20 PD/15 HC	Praat software; OpenSMILE toolkit (speech); STFT (speech and gait)	MDS-UPDRS SPEECH: 88 features (EGeMAPS); 2D speech - TFR; gait spectrograms; 1D handwriting signals	CNN; SVM	80/10/10 (training/Bayesian optimization/test); Kruskal-Wallis H-tests	**ACC**:92.3%; **AUC**:0.963 (Speech)	Diagnosis/ detection
[29]	2019	PC-GITA + SVD	50 PD/50 HC + 2000 samples (pathological speakers (not necessarily PD)/HC)	Randomly signal rolling; Random band-pass filter (low/high pass frequencies)	Frequency-based features from spectrograms	Modified ResNet-18 (ReLU)	90/10 (training/validation); 10-fold CV	**ACC**;91.7%; **F1-score**:0.92; **Prec**:92%; **AUC**:0.93 **Sen**:92%	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[30]	2019	2 databases + UCI	22 PD + 30 HC + 18 PD	Normalization (StandardScaler); PCA; Linear Regression; Random split (k-fold)	Praat script (Features: Jitter; Shimmer; Picth; Harmonicity; Gender)	Naïve–Bayes (NB);Perceptron algorithm; RF; SVM (RBF);NN	Optimization; LOO CV; Levene and Kruskal tests	**ACC**: 99.94% (RF); 92.38% (SVM); 91.10% (NN)	Diagnosis/ detection
[31]	2019	Hand PD Spiral + Hand PD Meander + Speech PD + Voice PD	105 PD/53 HC (handwritten) + 23 PD/8 HC (speech) + 20 PD/20 HC (voice)		Modified Grey Wolf Optimization algorithm (FS)	MGWO + ML (DT/RF/KNN (K=3))	70/30 (training/test)	**ACC**: 94.83%	Diagnosis/ detection
[32]	2020	UCI + PC-GITA	25 PD/20 HC + 25 PD/20 HC	Downsampling	Acoustic feature: (27 features); MFCC; EMD (IMF features:(Energy; Entropy; Spectral entropy; Intrinsic mode functioncepstral coefficient (IMFCC); Statistical features))	SVM; RF	80/20 (training/test); 10-fold CV; LOOCV; LOSO test; Cross-database evaluation	**ACC**: 98% (SVM)/100% (RF); **AUC**: 1.00 (RF)/0.99 (SVM); **Sen**: 100% (SVM); **Spec**: 92% (SVM)	Diagnosis/ detection
[33]	2020	Parkinson's UI	23 PD/8 HC	Praat software; Normalization	Traditional measures: Praat software (short-term autocorrelation); Nonstandard measures: (Time Series Analysis; DFA; normalization)	ANN (Levenberg-Marquardt training algorithm); KNN	70/25/5 (training/test/-validation) (ANN); 70/30 (training/test) (KNN); Overall accuracy	**ACC**: 96.7% (ANN)/79% (KNN, k=1)	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[19]	2020	UCI	23 PD/8 HC (139/56 (training/test) sustained vocal phonations)	Standardization; Zero score and Min-max normalization	Features: Frequency, jitter, shimmer, voice tone (HNR, NHR) Recurrent Period Density Entropy, Pitch Period Entropy	Naïve Bayes; LR; K-NN; RF	Performance metrics	**ACC**: KNN: 90.2% (K=5); **F1-score**: RF: 0.895; **Prec**: LR: 93%; **Sen**: RF: 92.8%	Diagnosis/ detection
[34]	2020	PC-GITA	50 PD/50 HC	Spectrograms	Deep learning convolution model; Alexnet; Handcrafted feature-based model (RMS; spectrum; zero crossing; pitch; entropy; MFCC; statistical features)	Transfer learning method (Alexnet); SVM; RF; MP; ANN	5-fold CV	**ACC**: (1) MP:99.7%/(2) RF:83.6%	Diagnosis/ detection
[35]	2020	Collected Dataset (AC+SP)	197 AC/198 SP (HC+PD)	80/20 (training/test); Feature Extraction	Praat software: Voice (V) and Unvoiced (UV) parts; Phonation and Speech; Dimensionality Reduction (Low Variance Filter	KNN (K=1); SVM	Performance metrics	**ACC**: AC: 86.52%/SP: 84.14%; **Prec**: AC: 86.52% /SP: 73.78%; **AUC**:AC: 0.8436/SP: 0.7815; **Sen**: AC: 85.43%/SP: 82.48% (KNN)	Diagnosis/ detection
[36]	2020	Audio database (Medical University of Warsaw)	22 PD/22 HC	Overlapping	LFCC analysis: homomorphic technique; MFCC: linear mapping of low frequencies and logarithmic mapping of high frequencies; GTCC: band-pass filter; Segmentation: $\Delta and \Delta\Delta parameters$	LDA; K-NN (k=2)	11-fold CV; linear cepstral technique ($\Delta and \Delta\Delta$)	**ACC**: LDA-MFCC:86.4%/ K-NN-MFCC: 83%; **Sen**:$\Delta + \Delta\Delta$: $80\%/LDA - GFCC$: $90.9\%/K-NN-MFCC$: 91.4%; **Spec** : $\Delta + \Delta\Delta$: $90\%/LDA - MFCC$: $90.9\%/K-NN-MFCC$: 77.3%	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
						Classification			
		Dataset	Participants	Preprocessing	Feature Extraction	Classifier	Validation	Results	Type
[37]	2020	Hungarian corpus	83 PD/33 HC	Normalization; VAD (Voice Activity Detection)	Forward-backward divergence segmentation (FBDS); Transient-stationary segmentation (TSS); Features: Length of segments, Speech rate and Speech rate changes	C-SVC+RBF	LOSOCV	**ACC**: TSS/FBDS:76%/73%; **F1-score**: TSS/FBDS:0.72/0.69; **Sen**: TSS/FBDS:61%/60%; **Spec**: TSS/FBDS:SYL-91%/85%	Diagnosis/ detection
[38]	2020	PC-GITA	50 PD/50 HC	Short Time Fourier Transformation (STFT)*; Continuous Wavelet Transformation (CWT)**	Spectrogram*; Scalogram**;Time-frequency features	Stacked auto-encoder DNN with a Softmax classifier	80/20; hold-out validation	**ACC**: STFT: 87%/CWT: 82%	Diagnosis/ detection
[39]	2020	UCI	23 PD/8 HC	Feature selection (Univariate Selection, Recuesive feature elimination, Feature importance)	22 phonetic features (Shimmer:APQ3, spread2, D2, Shimmer:APQ5, DFA, RPDE, PPE, spread1, spread2, Jitter:DDP MDVP:APQ, HNR, NHR, MDVP:Shimmer)	Classification and Regression Trees (CART); SVM; ANN	Performance metrics	**ACC**: (before FS/after FS) CART: 85.23%/90.76%/ SVM: 79.98%/93.84%/ ANN: 80.25%/91.54%	Diagnosis/ detection
[40]	2020	(1) MMPD + two noise data sets: (2) 4 types of noise; (3) 3 types of noise (a subset of NOISEX-92)	(1) 400 PD/400 HC/8000 random samples/20 outlier samples	Segmentation; Noise reduction, Dereverberation, Declipping algorithms	13 PLP coefficients (standard deviation of fundamental frequency, jitter, shimmer, HNR, glottal-tonoise excitation ratio, articulation rate, frequencies of formants); noise; reverberation; nonlinear distortion	Gaussian mixture models (GMMs) are fitted to the frames of the voice recordings parametrized by PLP coefficients	5-fold CV	**AUC**:0.95	Diagnosis/ detection (evaluate the impact of signal degradation and enhancement on PD detection performance)

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[41]	2021	Recordings at Hospital of Rome and Elderly's Center	33 PD/18 HC	Auto-cutting algorithm (BP filter; Down Sampling; Segmentation; TKEO Operator; Sum; Smoothing; Compressor; Amp Treshold; Time Treshold)	Standard Measures (Normalized Correlation Function, HNR (Autocorrelation Function)); Nonstandard Measures (RPDE; DFA); Cepstral Measures (CPPS; MFCC)	SVM (SMO; Matlab SVM);ANN (MLP); KNN (RSL, IBk)	75/25; 10-fold CV	**ACC**:97.3%/95.4%/ 87.6%/79%/91.2%; **F1-score**:0.97/ 0.94/ 0.84/ 0.68/ 0.90; **AUC**:0.97/ 0.94/ 0.91/ 0.74/ 0.95 (RSL/Ibk/MLP/SMO/-Matlab SVM)	Diagnosis/ detection
[42]	2021	(1) LSVT voice rehabilitation dataset + (2) Sakar dataset	14 PD + 20 PD/20 HC			New deep dual-side learning ensemble model	LOSO CV	**ACC**: (1) 98.49%/ (2) 99.67%; **Sen**: (1) 98.4%/ (2) 99.35%; **Spec**: (1) 99.1%/ (2) 99.7%	Diagnosis/ detection
[43]	2021	(1) NewHandPD (handwriting) + (2) PC-GITA (voice)	31 PD/35 HC + 50 PD/50 HC	(1) Re-scale; Argumentation (flipping, exposure, brightness); (2) Resampled to 16kHz; Converted to a Mel fbank spectrogram using a 25 ms Hamming window every 10 ms	Spectrogram	(1) Vision Transformer algorithm; (2) Audio Spectrogram Transformer Algorithm	80/20 (training/validation); 5-fold CV	**ACC**: (1) 92.37%/ (2) 87.5%	Diagnosis/ detection
[44]	2021	Collected database (Spain)	36 PD	Data augmentation	CPP, D2, RPDE, MFCC5	MCMC sampling algorithm		**ACC**: 71.43%	Prognosis

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[45]	2021	MDVR-KCL	16 PD/21 HC	ParseImouth library (Python)	Jitter class; Shimmer class; HNR	Decision Tree; Resnet50 NN model	Performance metrics (Confusion Matrix) (1000 iterations)	**ACC**: DT:61.2%/ RN50:97.3% **AUC**: DT:0.5278/ RN50:0.7232	Diagnosis/ detection
[46]	2021	UCI	23 PD/8 HC	Feature Selection (FS)	Correlation Heat map (relationship between features)	KNN; SVM; RF; NB; LR; Meta Classifier Model; Bagging Model; DT; Adaboost; Gradient boosting	Performance metrics (Confusion Matrix)	(best) **ACC**: (Train/Test) RF:98.53%/93.22% **F1-score**: MCM:0.9655; **AUC**: KNN:0.9455; **Sen**: SVM/GB/DT/BM:100%	Diagnosis/ detection
[47]	2021	Voice database, Faculty of Medicine, Istanbul University	188 PD/64 HC	Normalization (min-max)	SMOTE	(Stacking classifier and Voting classifier technique) AdaBoost; Extra Trees classifier; DT	80/20 (train/validation); 10-fold CV	**ACC**: Stacking classifier:92.2%/ Voting classifier: 83.57%	Diagnosis/ detection
[48]	2021	UCI (1) Acoustic dataset + (2) Parkinson's dataset	23 PD/8 HC + 40 PD/40 HC	Feature selection	(1) MDVP, Jitter, Shimmer, Harmonics Ratio, Entropy, frequency variation, PPE; (2) Pitch and Amplitude local perturbation measures, noise features, Spectral envelope and nonlinear measures	CNN (ReLU); ANN (Levenberg-Marquardt (LM))	85/15 (train/-validation); Confusion matrix	**ACC**:((1)/(2)) ANN: 82.76%/77.22% CNN: 93.10%/88.89%	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[49]	2021	Collected Dataset "De Novo PD and Healthy subject's data"	51 PD/51 HC	(FS) Filter (forward greedy step-wise filter with CFS evaluator); Ranker (Correlation, Information Gain, and Gain Ratio)	6373 Acoustic Statistical Parameters (Spectral, Prosodic, Voice quality and Cepstral)	Naive-Bayes; SVM-SMO; MLP	10-fold CV; Statistical Validation (Iman and Davenport test, Wilcoxon's test, Nemenyi's test)	**ACC**: NB:94.34%/ SVM-SMO:93.806%/ MLP:93.224%	Diagnosis/ detection
[50]	2021	2 voice datasets, Faculty of Medicine, Istanbul University	18 PD/20 HC + 20 PD/20 HC	RELIEF algorithm (FS); Normalization	Praat software (Time frequency domaine: Jitter; Shimmer; HNR; Pitch; RPDE; DFA; PPE; CD: MFCC; PLP; RASTA-PLP); Voice Analysis Toolbox (Cepstral domaine; Wavelet Transform)	SVM (RBF and Polynomial); KNN (k=5)	train/validation; Performance metrics	(KNN-SVM'poly'-SVM'RBF') **ACC**: TFD: 60%-70%-70%/ CD:45%-80%-65%/ WT:50%-45%-45%; **Sen**: TFD:46%-100%-92%/ CD:31%-69%-62%/ WT:54%-23%-31%; **Spec**: TFD: 86%-68%-71%/ CD:67%-100%-80%/ WT:64%-75%-67%	Diagnosis/ detection
[51]	2021	GYENNO SCIENCE Parkinson Disease Research Center	30 PD/15 HC	PCA; DL Models: librosa	NeuroSpeech software (phonation (P); articulation features (A)); DL Models: linear (STFT) spectrogram; Mel-scaled STFT spectrogram; Constant-Q transform spectrogram	ML models (DT, MLP, KNN, Gaussian Naive Bayes, SVM); DL models (CNN, RNN); Bidirectional LSTMs model	(1) 10-fold CV; (2) training/testing sets	(best) **ACC**: (1) RNN (B.LSTM): 84.29%/ (2) B.LSTM: 75.56%; **F1-score**: (1) RNN (B.LSTM): 0.8852/ (2) B.LSTM: 0.807; **MCC**: (1) RNN (B.LSTM): 0.6603/ (2) B.LSTM: 0.4811; **Sen**:(1) RNN (B.LSTM): 87.34%/ (2) B.LSTM: 85.19%; **Spec**: (1) DL (Mel-Spect): 92.22%/ (2) SVM (poly): 97.67%	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
						Classification			
		Dataset	Participants	Preprocessing	Feature Extraction	Classifier	Validation	Results	Type
[52]	2021	(1) PC-GITA + (2) Parkinson's foundation of Medellín	50 PD/50 HC + 20 PD/20 HC	(1) Downsampling (linear interpolation)	STFT (Hanning windows with 512 samples with an overlap of 256 samples); Non-negative matrix factorization; TF-based feature (Renyi entropy; 17 features); Baseline/Acoustic features (jitter, shimmer, GNE, VFER, HNR, NHR, GQ); MFCC	SVM (classification); SVR (estimate dysarthria level)	Kruskal-Wallis tests; LOSO-CV (train); 80/20 (train/test) (80x - 80 models)	**ACC**: (1) TF: 97% B/A: 62% MFCC: 61% / (2) TF: 71% B/A: 55% MFCC: 57%; **F1-score**: (1) TF: 0.98 B/A: 0.57 MFCC: 0.57; **MCC**: (1) TF: 0.96 B/A: 0.26 MFCC: 0.54; **AUC**: TF: 0.99 B/A: 0.83 MFCC: 0.87; **Sen**: (2) TF: 66% B/A: 67% MFCC: 56%; **Spec**: (2) TF: 76% B/A: 52% MFCC: 63%	Diagnosis/ detection
[53]	2021	MDVR-KCL + Collected Dataset	16 PD/21 HC + 20 HC	Segmentation; Noise reduction; Resampling (Audacity software)	RSSD signal decomposition technique (TQWT technique; SALSA); Time-Frequency analysis (Hamming windows; 50% overlap); PSD calculated (converted into a signal image)	CNN model (best result: 64 neurons (dense layer))	(1) 70/30 (training/-validation); (2) 50/25/25 (training/-validation/testing); Kolgomorov-Smirnov test	**ACC**:98.12% (1)/87.50% (2); **F1-score**: 0.97 (1)/ 0.76 (2); **Prec**: 96% (1)/ 100% (2); **AUC**: 0.88 (2); **Sen**: 97% (1)/ 62% (2)	Diagnosis/ detection
[54]	2021	mPower app	18660 HC/8442 PD (train); 4628 HC/2147PD (validation); 200HC/200HC (test)	Downsampling and reshaping frequency window	2D representation of sounds, such as STFT-DCT spectrogram	CNN models (DenseNet-161, ResNet-50, SqueezeNet1-1)	Performance metrics	**ACC**: DN161: 91.17%/ RN50: 90.76%/ SN1-1: 81.85%; **Prec**: DN161: 89.25%/ RN50: 88.87%/ SN1-1: 73.68%; **Sen**: DN161:91.5%/ RN50: 90%/ SN1-1: 74%	Diagnosis/ detection
[55]	2021	Mobile Parkinson Disease Study	700 PD/700 HC	Normalization	Shimmer, jitter, Min + max pitch, minimum tone, HNR, nº pulses, fundamental frequency, MFCCs, Dimensionality reduction techniques: HCF+PCA	SVM; RF; LR; MLP	Assessment metrics (confusion matrix)	**ACC**: HCF/PCA MLP: 86%/ RF: 82%/ SVM: 88%/ LR: 80; **AUC**: MLP:0.904/ RF: 0.903/ SVM: 0.919/ LR - 0.844	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[56]	2021	PC-GITA	50 PD/50 HC	Frame process (25 ms); Hamming window and 50% overbinning	VMD method; Mel-Spectrogram extraction	ResNet-LSTM network (Softmax)	10-fold CV	**ACC**: RN18-LSMT: 96.12%/ RN50-LSMT: 95.96%/ RN101-LSMT: 97.2%; **Prec**: RN18-LSMT:98.14%/ RN50-LSMT: 96.7%/ RN101-LSMT: 97.8%; **Sen**: RN18-LSMT:96.74%/ RN50-LSMT: 96.21%/ RN101-LSMT:95.78%	Diagnosis/ detection
[57]	2021	1 Dataset	20 PD/14 HC	Pre-emphasizing, framing, windowing	Genetic Algorithm (feature selection); Discrete wavelet transform; Formant frequencies (LPC); Time-domain energy and ZCR; MFCC; Shannon entropy decompositional features	SVM with 10 fold CV	train/test; LOSO validation	**ACC**:DWT+MFCC+SVM: 91.18%/ DWT+MFCC: 86.84%	Diagnosis/ detection
[58]	2021	PC-GITA	50 PD/50 HC	Speech: down-sampled; Features: normalization mean, standard deviation, min, max, kurtosis, skewness (Baseline features), downgrading, WLP analysis (Glottal features)	Traditional pipeline approach (Baseline features (Articulation, Phonation, Prosody) (NeuroSpeech toolkit) + glottal features (IAIF and QCP analysis) (APARAT Toolbox)); End-to-end architecture (raw time-domain waveform (speech signal, voice source (ZFF)) + glottal features)	SVM (TPA); CNN+MLP (EEA)	10-fold CV; 90/10 (train/test) Train:90/10 (train/validation)	TPA/EEA (best) **ACC**: B+G(QCP): 67.93%/ G(QCP): 68.56%; **Sen**: B+G(QCP): 69.71%/ G(QCP): 63.4%; **Spec**: G(QCP): 70.43%/ Voice(ZFF): 78.86%	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[59]	2021	UCI	23 PD/8 HC	Normalization; Multi-agent Feature Filter (MAFT) Algorithm (feature selection and evaluation)	Original DB: 22 features; Filtered DB: 14 features (Original processed to obtain a new filtered database)	Hybrid Model (binary CNN + 3 feature selection algorithms (GA, AO, and M-BGD))	10-fold CV (100 tests); 90/10 (train/test); Standard evaluation metrics	(Original/Filtered DB) **ACC**:93.7%/96.8%; **F1-score**:0.922/0.943: **Prec**: 92.4%/93.4%; **AUC**: 0.91/0.93; **Sen**: 91.5%/92.2%	Diagnosis/ detection
[60]	2021	UCI	124 PD/64 HC (564 PD + 192 HC rows)	Adaptive Crow Search Algorithm (ACSA) (feature selection); Splitting into training and testing sets	Wrapper, embedded, filter methods	SVM; LDA; GBM; RF; NB; LR	K-fold mechanism; Standard metrics	(Nº features: 754/36/7) **ACC**: LDA: 72.2%/75.4%/96.4%; **Sen**: LDA: 74.8%/75.4%/96.4%; **Spec**: LDA: 72.3%/77.2%/98.5%	Diagnosis/ detection
[61]	2021	Hand PD meander and spiral (1) + PD speech (2) + PD acoustic (3) + Parkinson Disease Classification Data Set (speech 2) (4)	(1) 53 HC/105 PD (632 images) + (2) 20 HC/20 PD (26 samples) + (3) 40 HC/40 PD (45 features) + (4) 64 HC + 188 PD	Normalization	Feature weight (Class separability; Pearson correlation coefficient; ReliefF)	Multi-filter hybrid feature selection algorithm model based on discrete artificial bee colony (MFABC)	Split into train/test (70/30);	**ACC**: Multiple classifiers Hmeander: 96.83%/ Hspiral: 97.62%/ Speech: 100%/ Acoustic: 95.83%/ Speech 2: 91.97%	Diagnosis/ detection
[62]	2022	UCI	23 PD/8 HC	Standardization; PCA; Feature scaling train/test	22 extracted features; Feature vector matrix	SVM; RF; KNN	Performance measurements	**ACC**: SVM: 81.7%/ RF: 95%/ KNN: 81.7%; **F1-score**: SVM: 0.925/ RF: 0.969/ KNN: 0.921; **Prec**: SVM: 86.5%/ RF: 96.9%/ KNN: 93.5%; **AUC**: SVM: 0.919/ RF: 0.977/ KNN: 0.937; **Sen**: SVM: 100%/ RF: 96.9%/ KNN:90.6%	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[63]	2022	CLP data + PC-GITA	135 CLP/58 HC + 50 PD/50 HC	Augmentation output images (Row-Resampling; Adding noise; Clipping; Speed Up; Harmonic distortion; Gain; Rand time shift; Sound mix; dynamic Range; Pitch shift; Lowpass filter)	SpecAugment data augmentation; Spectrogram (126-time steps and 128 Mel-scale frequency bins)	IPWOA-DCNNs	10-fold CV; 80/10/10 (train/validation/test)	**ACC**:95.77%; **F1-score**:0.9544; **Prec**:96.33%; **Sen**:93.33%; **Spec**:93.55%	Diagnosis/ detection
[64]	2022	GYENNO SCIENCE database + PC-GITA	30 PD/15 HC + 50 PD/50 HC	overlapping segments on the log Mel-spectrograms; time-distributed 2D-CNN blocks	(1) Articulation features (NeuroSpeech); (2) Phonation features (NeuroSpeech); (3) subset of Surfboard components	2D-CNN + 1D-CNN	(1) 60/40 (2) 80/20 (train(validation)/test); 10-fold CV	**ACC**:(1) 81.6%/ (2) 92%; **F1-score**:(1) 0.8766/ (2) 0.9104; **MCC**:(1) 0.5847/(2) 0.856; **Sen**:(1) 79.17%/ (2) 91.21%; **Spec**:(1) 98.33%/ (2) 100%	Diagnosis/ detection
[65]	2022	PC-GITA	50 PD/50 HC	augmentations (Time shifting; Band-pass filter; Pitch change; Slow-down; Speed-up; Colored noise)	STFT (log-frequency power spectrograms)	CNN (optimization (binary cross-entropy lossfunction, stochastic gradient descent (SGD) optimizer algorithm); evaluate the efficiency (Xception network model))	10-fold CV	(Vowels/Words) **ACC**: 82.67%/82.12%; **AUC**: 0.8695/0.8835; **Sen**: 83.33%/88.16%; **Spec**:81.84%/76.1%	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[66]	2022	PC-GITA (training/-validation) + Saarbruecken Voice Database (SVD) (training) + Vowels dataset (training)	50 PD/50 HC + 1355 D/687 HC	STFT; Gaussian blurring to the spectrogram log-frequency spectrograms resized to 224 × 224 pixels	extrated for CNN (not require supervised feature extraction)	MFT CNN	90/10 (train/test); 10-fold CV	**ACC**:99%; **AUC**: 0.896; **Sen**: 86.2%; **Spec**:93.3%	Diagnosis/ detection
[67]	2022	CLP data + PC-GITA	135 CLP/58 HC + 50 PD/50 HC	Augmentation process (Row-Resampling; Gaussian noise; Clipping; Speed Up; Harmonic distortion; Gain; Rand time shift; Sound mix; dynamic Range; Pitch shift; Lowpass filter)	SpecAugment data augmentation Spectrogram (126-time steps and 128 Mel-scale frequency bins)	IPChOA-DCNN model	10-fold CV; 80/10/10 (train/development/test)	**ACC**:96.22%; **F1-score**:0.9623; **Prec**:97.01%; **Sen**:94.18%; **Spec**:94.25%	Diagnosis/ detection
[68]	2022	PC-GITA	50 PD/50 HC	hamming window	13 coefficients of MFCCs and SFFCCs;SDC features	LSTM and BiLSTM models (LSTM; BiLSTM; MHA-BiLSTM)	Performance measurements	(LSTM/BiLSTM/MHA-BiLSTM) **ACC**:MFCCs+SDC: 77.71%/81.5%/85.02%; **F1-score**: SFFCCs+SDC: 0.807/0.832/0.827	Diagnosis/ detection
[69]	2022	Kaggle	195 recordings	Min-Max Scaler; Split train/test (79/21)	22 features	LightGBM Classifier	Performance measurements	**ACC**:97.56%; **F1-score**:0.985	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[70]	2022	UCI	23 PD/8 HC	Splitting into train/test (70/30)	22 features (Jitter; Shimmer; HNR;NHR;DFA;D2;PPE; MDVP; Spread)	LR; KNN; SVM; RF; Adaptive Boosting; Stacking	Performance measurements	(LR/KNN/SVM/Stacking /RF/AB) **ACC**: 86%/92%/95%/95%/ 91%/89%; **F1-score**: 0.91/0.95/0.97/0.97/ 0.95/0.94; **Prec**: 91%/90%/94%/94%/92%/92%; **Sen**: 91%/100%/100%/100%/ 98%/96%	Diagnosis/ detection
[71]	2022	iPrognosis app (Gdata (4 lang)/Sdata (3 lang))	106 PD/392 HC (train) + 39 PD/24 HC (test)	Standardization process resampling; digital zero removal; Direct Current (DC) offset removal; normalization	auto-correlation combined with peak pruning (pitch); discrete cosine transform (13 MFCCs, 22 BBE)	(1) Lasso, Ridge, Gini Impurity, ANOVA + LSVM, LR, RF + median aggregation; (2) median aggregation + feature selection (1) + LSVM, LR, RF; (3) feature selection (1) + Multiple Instance Learning (MIL)	LOSO CV (GData)	**AUC**: 0.69/0.68/ 0.63/0.83 (English/- Greek/German/ Portuguese-speaking)	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[72]	2022	UCI	23 PD/8 HC (195 recordings (147 PD + 48 HC))	75/25 (train/test)	22 characteristic attributes for dysphonia (Jitter; Shimmer; HNR;NHR;DFA;D2;PPE; MDVP; Spread, Status)	Optimal XGBoost (randomized search and grid search)	Cross-validation; Performance measurements (Confusion matrix); Model comparison	**ACC**:96%; **F1-score**:0.97; **Prec**:100%; **AUC**:0.97; **Sen**:95%	Diagnosis/ detection
[73]	2022	UCI	188 PD (recordings)	Standardization; Mapping feature importance (xgboost feature significant); split data train/test	vocal fold, TQWT Features, Wavelet Features, MFCC, baseline characteristics	XGBoost	Performance measurements (Confusion matrix)	**ACC**: All Features:84.80%/ (1) excluded (ddaShimmer,locShimmer) 85.60%/ (2) excluded (locDbShimmer, meanNoiseToHarmHarmoniciy,ppq5jitter, apq5Shimmer, ddpJitter, rapJitter,PPE): 84.40%	Diagnosis/ detection
[74]	2022	UCI	23 PD/8 HC (195 recordings)	Data cleaning; Missing values handling; Categorical Variables transformation; Oversampling (SMOTE)	24 clinical characteristics (MDVP, Jitter, RAP, DDP, PPQ, shimmer,APQ, APQ3, APQ5, DDA, NHR, HNR, RPDE, D2. DFA. spread 1, spread 2, PPE, Health status of the subject)	LightGBM; RF; XGBoost, AdaBoost, Bagging, DT, LR, SVM, KNN, NB	80/20 (train/test); hold-out validation technique	(LightGBM) **ACC**:95%; **F1-score**:0.90; **AUC**:0.96; **Sen**:100%; **Spec**:93.33%	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[22]	2022	Neurology Outpatient Clinic and Department of Neurology at John Paul II Hospital	27 PD (5x/perPat (off state; 30min; 60 min; 180min after medication))	Pearson and Spearman Correlation	Signal parameterization; Phonatory and articulatory analysis	Multiple linear regression (MLR); $\varepsilon - SVR; RF$	10-fold CV (validate the algorithms); 5-fold CV; 80/10/10 (train/validation/test)	**MAE**(UPDRS-III Prediction vowel /a/) RF: 1.8530	Prognostic
[75]	2022	UCI	40 PD/40 HC	70/30 (train/validation+test); heatmap	44 acoustic features, amplitude local perturbation measures, noise features	FFNN	cross-entropy; confusion matrix	**ACC**:85%; **F1-score**:0.8435; **MCC**:0.7024; **Prec**:80.83%; **Sen**:88.18%; **Spec**:82.31%	Diagnosis/ detection
[76]	2022	UCI	23 PD/8 HC (195 recordings)	Data cleansing L1 and L2 regularizations; Adam (adaptive estimating, optimization, a stochastic gradient descent approach)	Feature Scaling (MinMax scaler method); Dense and activation layers	DNN Model (5 layers)	70/30 (train/test); feature correlation heat map	**ACC**: 94.87% (validation)	Diagnosis/ detection
[77]	2022	1 database	40 PD/40 HC (45 acoustic features)		LightGBM + SVM + Spearman correlation coefficient analysis (feature importance)	LightGBM	5-fold CV; 70/30 (train/test)	**ACC**: 83.23%; **Prec**: 86.84%; **AUC**: 0.87; **Sen**: 78.57%; **Spec**: 87.95%	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[78]	2023	1 dataset (train/test)	20 PD/20 HC (26 samples) - train; 14 PD/14 HC (1208 recordings) - test	80/20 (train/validation)	Mel-frequency cepstral coefficients extration (39 MFCC); I-vectors extraction (GMM-UBM applied on the MFCC vectors)	(1) CNN features and MLP based system (TRAIN); (2) CNN features and SVM based system; (3) I-vectors and SVM based system	5-fold CV	**ACC**: (1) 60%/ (2) SVM-RBF: 75%/ (3) SVM-poly(200):97.68%; **F1-score**:(1) 0.5/ (2) SVM-RBF:0.65/ (3) SVM-poly(200): 0.94; **Prec**:(1) 52%/ (2) SVM-RBF:74%/ (3) SVM-poly(200):94%; **AUC**:(2) SVM-RBF:0.75/ (3) SVM-poly(200): 0.97; **Sen**: (1) 48%/ (2) SVM-RBF:0.62%/ (3) SVM-poly(200):96%; **Spec**:(1) 50%/ (2) SVM-RBF:65%/ (3) SVM-poly(200):93%	Diagnosis/ detection
[79]	2023	1 database	50 PD/50 HC	PCA (feature selection); MinMax scaling function (feature scale); 80/10/10 (train/val/test)	MFCC features; Voice Onset Time (VOI) Features; (29 features comprising 26-MFCCs, VOT, and two spectral moments)	fully-connected deep neural network (FC-DNN) (Adam+ReLu+binary-cross entropy (loss funtion), Tense Flow+Softmax (class))	10-fold CV	**ACC**:98%; **F1-score**:0.975; **MCC**:0.97; **Sen**:97%	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[80]	2023	1 database	42 PD (5875 voice measurements)	Imputed KNN method (check missing and imputed data); Data normalization	BALO feature selection approach (Non-parametric Wilcoxon significant test; cross-correlation analysis (var1:Jitter, var2:Jitter (RAP), var3:Jitter (PPQ5), var4:Shimmer (APQ3), var5:Shimmer (APQ1), var6:Shimmer (DDA), var7:NHR, var8:HNR, var9:RPDE, and var10:PPE))	BALO-DEELM	70/30 (train/test); Performance measures	(Sigmoid-60:UPDRS Motor/UPDRS Total) **ACC**:90.80%/91.23%; **MAE**:0.400/0.395	Prognostic
[81]	2023	PDO Dataset + PD Dataset	23 PD/8 HC + 188 PD/64 HC (756 recordings)	min-max normalization	SkipConNet (baseline feature, time–frequency features, MFCCs, wavelet transform-based features, vocal fold features, and TWQT features)	SkipConNet + RF	Confusion matrix	(PD/PDO) **ACC**:99.11%/98.30%; **F1-score**:0.99/0.97; **Prec**:99/99%; **AUC**:0.9877/0.9583; **Sen**:99%/96%; **Spec**:98.77%/95.83%	Diagnosis/ detection
[82]	2023	PSV dataset	26 recordings (PD+HC)	F-MDI, F-CORR, F-PER methods;"Pearson's Correlation"; 80/20 (train/test); normalization using the Min–Max scaling technique	28 features (Jitter; Shimmer; AC; NTH; NTN: Pitch; Standard deviation; etc)	NNB (hybrid k-NN + GB model)	Cross-validation; performance metrics	**ACC**:75.48%; **F1-score**:0.7506; **Prec**:75.63%; **AUC**:0.7492; **Sen**:74.92%	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[83]	2023	Neurology Outpatient and Department of Neurology at the John Paul II Hospital in Krakow, Poland	50 PD recordings (UPDRS-III determined)	Phoneme extraction; Segmentation; Normalization	Acoustic analysis (fundamental frequency, shimmer and jitter coefficients, 12 MFCC, energy, average power, zero-order, 1ª, 2ª and 3ª order moments, kurtosis, power factor, 1st, 2nd, 3rd formant's amplitude, 1st, 2nd, 3rd formant's frequency, signal to noise ratio, max, min, mean value and standard deviation)	SVM-RBF; $\varepsilon - SVR; GPR$	10-fold CV; 90/10 (train/test)	(vowel/consonants/ speech) **MAE**:$\varepsilon - SVR$: $10.27/10.31/10.23; GPR$: $9.59/10.27/9.61; SVM$: $8.52/11.26/9.77$	Prognostic
[84]	2023	PPMI database	23 PD/8 HC	Data wrangling	PCA (jitter, shimmer and MDVP of vowel phonations)	3 approachs (LR, SVM, RF, KNN) (1) Models (195 records + 22 features); (2) PCA applied (195 records + 5 features); (3) Imbalance removal(109 records + 22 features)	75/25 (train/test); Performance metrics (confusion matrix)	**ACC**: (1)RF:91.83%/ (2)SVM:91.75%/ (3)KNN:91.83%; **Prec**: (1)LR/SVM:100%/ (2)LR/RF/SVM:100%/ (3)KNN:95%; **AUC**: (1)RF/KNN:0.701/ (2)RF:0.818/ (3)KNN:0.883; **Sen**: (1)RF/KNN:86%/ (2)RF/KNN:90%/ (3)KNN:95%	Diagnosis/ detection
[85]	2023	KayPENTAX Model 4337	23 PD/8 HC	Normalization (min-max)	Fundamental frequency value; frequency amplitude change; Jitter; Shimmer; HNR; NHR; DFA (RPDE; D2; Spread 1 and 2; PPE)	RF (grid search method)	80/20 (train/test); Performance metrics (confusion matrix)	**ACC**:94.87%; **F1-score**:0.9359; **Prec**:96.43%; **AUC**:0.9889; **Sen**:0.9643	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[86]	2023	UCI	40 PD/40 HC	Histograms; Variable significance analysis (find key features)	44 acoustic characteristics (voice frequency, fundamental frequency, amplitude variation, nonlinear dynamical complexity, nonlinear basic frequency, noise-to-tonal component ratio in the voice)	AdaBoost; GB; Light GB; XGradient Boost	70/30 (train/test); Performance metrics (confusion matrix)	(XGB Best) **ACC**:87.39%; **F1-score**:0.8727; **Prec**:87.91%; **AUC**:0.8737 (XGB); **Sen**:87.39%	Diagnosis/ detection
[87]	2023	Parkinson's database	55 PD/64 HC	STFT - 135 grayscale spectrograms; Data augmentation technique - 1755 color spectrograms	Feature vectors obtained from an intermediate layer of the CNN	CNN (pre-trained models: AlexNet; VGG-16; SqueezeNet; Inception V3; ResNet-50); Hybrid CNN-ELM model ((1) grayscale - original; (2) color - original; (3) color - fragments; (4) color - fragments and original	10-fold CV; 70/10/20 (train/validation/test)	**ACC**: (1) ALX-ELM/RN5-CNN: 91.30% (2) ALX-ELM/ALX-CNN: 95.65% (3) ALX-CNN/RN5-ELM/VGG-ELM: 95.65 (4) ALX-CNN/IV3-CNN: 95.65	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[88]	2023	MDVR-KCL	394 PD/615 HC (samples)	Downsampled at 8kHz sampling rate; Sliding frame size set to 128 with a 75% overlap	STFT (Image of Running Speech)	DCNN (ReLU)	756/244 (train/test) recordings; Performance measurements	**ACC**: 99.48% (training)/ 79.10% (validation)	Diagnosis/ detection
[89]	2023	MDVR-KCL	16 PD/21 HC	VAD algorithm (Segmentation) (658 HC; 249 HY2; 140 HY3; 39 HY4)	WS transform (Scattering)	BO approach (Optimal parameters); SVM (OneVsAll); DT (GDI); Ensemble learning (AdaBoost); NB; KNN (k=1); MLP; DA (LDA)	80/20 (train+ validation/ test); Performance measurements; WMV (SVM, AdaBoost, KNN, MLP, LDA)	**Overall ACC**:98.62%; **AUC**: MLP: 0.98; (WMV) (HC/PDH2/PDH3/PDH4) **F1-score**: 0.9961/0.9898/ 0.9642/0.8888; **Prec**: 100/100/96.42/80%; **Sen**: 0.9924/0.98/ 0.9642/1.00; **Spec**: 1.00/1.00/ 0.9947/0.9963	Diagnosis/ detection + Prognostic
[90]	2023	PC-GITA	50 PD/50 HC	GA (FS and optimization)	WSST-based and TFRM-based	SVM and GBM	10-fold CV	**ACC**:86% (vowels (/a, i/))/ 95% (isolated words (/apto/))	Diagnosis/ detection
[91]	2023	1 dataset	37 PD (74 samples)	Data Aggregation and Missing Data Handling; Data Normalization	RF (FS) (Group: Motor Function; Fine-motor; Memory; Executive function; Gross-Motor; Multi-function; Speech (average time for a correct response, total correct, total generated, and number of missed words))	Hybrid RF/PCA (Correlation of Functions and Stages)			Prognostic

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[92]	2023	PC-GITA + ItalianPVS	50 PD/50 HC + 28 PD/22 HC	Image resized	SLT (TF spectrograms)	DNN (Relu+Softmax); (Inception-ResNetV2, ResNet50V2, VGG-16)	10-fold CV	(VGG-16) (PC-GITA/ItalianPVS) **ACC**:92%/96%; **F1-score**:0.93/0.96; **Prec**:95/99%; **Sen**:92%/93%; **Spec**:91%/99%	Diagnosis/ detection
[93]	2023	MDVR-KCL	37 PD/HC	Max-pooling, batch normalization and leaky ReLU activation functions	Auto-encoder	SOTA model under FL strategies (FedAvg; FedAvgM; FedAdam)	60/20/20 (training-validation-test); 10 fold CV	**ACC**: 61.49% (FedAdam (0.1/0.1))	Diagnosis/ detection
[17]	2023	UCI	23 PD/ 8 HC	Simple Imputer, Normalization	Feature selection	Hybrid Ensemble learning model (SVM+ XGBoost + LR + RF +DT)	75/25 (train/test+val)	**ACC**:85-98%; **F1-score**:0.72-0.94; **Sen**: 80.5-96%	Diagnosis/ detection
[94]	2023	UCI	23 PD/ 8 HC		Pearson's correlation coefficient; Feature selection	DT + RF + +LR + SVM + NB (Comparison of MLs)	70/30 (train/test); 5-fold CV; Stratified 5-fold CV	RF (70-30/k-f/Strat k-f) **ACC**:95.42%/91.79%/ 91.28%	Diagnosis/ detection
[95]	2023	UCI	70PD/15HC	Heatmap; Feature engineering	30 features	SVM, LR, KNN	train/test	**ACC**: 60-100%; **Prec**: 40-50%; **Sen**: 57-100%; **F1-score**: 0.47-0.67	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[96]	2023	PD DB	160 PD/266 HC + 23PD/8HC	Undersampling; SMOTE; Random oversampling	Information Gain (IG) ranking; Correlation-based feature selection (CFS); mRMR; Mel-spectrograms; 6 audio data enhancement methods (Time stretching; Pitch shifting; Noise addition; Room simulation; Time masking; Frequency masking)	CNN Hybrid model (LSTM and GRU models)	Split data	**ACC**:93-96% **Prec**: 84-97%; **Sen**: 72-98%; **F1-score**: 0.78-0.96	Diagnosis/ detection
[97]	2023	PC-GITA + 2 datasets	50PD/50HC + 27PD/27HC + 27PD	VAD removal; Normalization	MFCC; Mel-spectrogram	ViT (Vision Transformer) (DNN model)	5-fold; training/validation split with 4:1 ratio; Confusion matrix	**F1-score**: 0.78; **Sen**: 78%; **Prec**: 77%; **AUC**: 0.83	Diagnosis/ detection
[98]	2023	PC-GITA	50 PD/ 50 HC	Downsampled	2D recurrences plots	CNN	10-fold CV	**ACC**: 87%; **Prec**: 92%; **Sen**: 84%; **Spec**: 91%	Diagnosis/ detection
[99]	2023	UCI + MDVR-KCL	23PD/8HC	PCA; Oversampling	Praat software (MDVP, NHR, HNR, Jitter, Shimmer, VTI, ATRI, DFA); Spectrograms	SVM, CNN, DT, NB, KNN	Confusion matrix	**ACC**: 53.27-82.24%; **Prec**: 51-89%; **Sen**: 22-92%; **F1-score**: 0.32-0.83 (Best:DT; Worst:NB)	Diagnosis/ detection
[100]	2023	UCI + Spiral dataset	195 voice samples + 512HC/250PD	Regression		Voice - XGBoost; Spiral - RF	Confusion matrix	**ACC**: XGBoost - 94%; **ACC**: RF - 91%	Diagnosis/ detection

Table 1: Related works (Continued)

Study	Year	Data		Algorithm				Performance	
		Dataset	Participants	Preprocessing	Feature Extraction	Classification		Results	Type
						Classifier	Validation		
[101]	2023	UCI	23PD/8HC	Standardization	Heatmap; Box Plot; Histogram plot	SVM, DT, RF, LT, Gboost, XGBoost	Split data; Confusion matrix	**ACC**:85-87%	Diagnosis/ detection
[102]	2023	UCI	188PD/64HC	Resampling	Vocal fold, TF, MFCCs, WT based and TWQT	LSTM	90(20)/10 train(val)/test	**ACC**: 93%; **Prec**: 96%; **Sen**: 90%; **F1-score**: 0.93	Diagnosis/ detection
[103]	2023	PD DB	40PD/40HC	Denoising (Weiner filter)	13 MFCCs	DT, KNN, RF, SVM	70/30 train/test	**ACC**: 95.8% (RF)	Diagnosis/ detection
[104]	2023	PD DB	23PD/9HC	Normalization; P-value test and Heatmap	MDVP, Jitter, Shimmer, Nonlinear measures	LR, KNN, DT, SVM, NB, Boosting algorithms, RF, ANN	k-fold CV	**ACC**: >85%	Diagnosis/ detection
[105]	2023	MDVR-KCL + PC-GITA	16PD/21HC + 50PD/50HC	Audacity audio editor software; 3s non-overlapping chunks; Normalization	INTERSPEECH 2010 paralinguistic challenge (IS10) feature set	SVM + RF	5/10-fold CV; 80/20 training/testing	(IS10 feature set, Proposed-NSRC) **Prec**: 81%; **F1-score**: 0.8371; **MCC**:0.47; **ACC**: 78.88%; (Combined feature set, Proposed-NSRC) **Sen**: 89.84%; **F1-score**: 0.8614; **ACC**: 82.46% ; **MCC**:0.53: (Text reading task, Proposed-NSRC)**F1-score**: 0.8103; **Prec**: 79.66%; **ACC**: 83.08%; **MCC**:0.57	Diagnosis/ detection

2.2.3 Discussion of the Systematic Literature Review

2.2.3.1 Dataset

According to Table 1, the reported studies use one or more databases for the detection and prognosis of PD, which may have been collected by the authors or come from available repositories with audio recordings of the disease. Of the databases from repositories, three datasets stand out: UCI database (UC Irvine Machine Learning Repository) with 29 articles of 87 (available online: `https://archive.ics.uci.edu/`), PC-GITA database [106] with 20 articles, and MDVR-KCL [107] with 7 articles.

UCI database is a repository with 663 databases on different pathologies, three of which stand out in Table 1, mentioned above. These are the Parkinson's Speech Dataset with Multiple Types of Sound Recording, the Oxford Parkinson's Disease Detection Dataset, and the Parkinson's Disease Classification.

Parkinson's Speech Dataset with Multiple Types of Sound Recording (available online: `https://archive.ics.uci.edu/dataset/301/parkinson+speech+dataset+with+multiple+types+of+sound+recordings`), recorded at the Department of Neurology in Cerrahpasa Faculty of Medicine, Istanbul University, is a training and testing dataset consisting of sound recordings from 40 people, 20 of whom have PD and the remaining 20 are healthy control, and another group of 28 PD patients. The training part consist of 26 voice recordings of each subject, for the first group, and is made up of voice samples of sustained vowels (/a/, /o/, /u/), short sentences, words and numbers from 1 to 10, from that group, the linear and time-frequency based features were extracted from each voice samples, as well the UPDRS score of each patient. The test part is made up of 168 voice recordings from the 28 patients with PD, with voice samples of sustained vowels / a / and / o /.

Oxford Parkinson's Disease Detection Dataset (available online: `https://archive.ics.uci.edu/dataset/174/parkinsons`) is composed with voice measurements, such as average, maximum and minimum vocal fundamental frequency, several measures of variation in fundamental frequency and in amplitude, two measures of ratio of noise to tonal components in the voice, the status of the subject, two nonlinear dynamical complexity measures, the signal fractal scaling exponent and three nonlinear measures of fundamental frequency variation of 23 PD patients and 8 healthy people.

Parkinson's Disease Classification (available online: `https://archive.ics.uci.edu/dataset/470/parkinson+s+disease+classification`) has three sustained phonation recordings of the vowel /a/ from each of the 252 subjects, including 188 patients with PD and 64 people in the healthy control.

The second most used database was PC-GITA [106]. This dataset consists of speech recordings, such as vowels and sustained vowels, words, complex and simple sentences, and monologs, of 100 Colombian Spanish native speakers, of which 50 are PD patients and 50 are healthy people. In this database, age, UPDRS score and Hoehn & Yard scale (H&Y) are provided.

Mobile Device Voice Recordings at King's College London (MDVR-KCL) dataset was the

next most refered in table 1. MDVR-KCL (available online: https://zenodo.org/records/2867216) was recorded at King's College London (KCL) Hospital, Denmark Hill, Brixton, London, being composed of 42 high quality voice recordings in the HC group and 31 in the PD group, recordings that are divided into tasks such as reading texts and spontaneous conversations. The scale of Hoehn & Yahr (H&Y) and UPDRS score is also provided.

2.2.3.2 Preprocessing

The vast majority of the databases used in the studies mentioned above refer to the care with which the voice and speech files were recorded, in order to obtain the best signal quality with the lowest possible level of interference, for example, the existence of excessive noise, something that may influence the subsequent analysis and evaluation of the signal.
In this case, taking Table 1 into account, the most commonly used preprocessing methods were normalization, with 30 articles, where, as the name suggests, the signal is normalized, i.e. the signals are transformed so that all lie within the same range, as for example in [79].
The use of short-time fourier transform (STFT), with 9 studies, segmentation, with 8 articles, and principal component analysis (PCA), with 8 studies, it's also common, with the STFT being used to determine time-frequency components. Another method used is Praat software [108], with 8 articles of 87, this free software was created to analyse sound signals, making it possible to carry out phonetic measurement and analysis, applying filters, sampling and resampling methods, among others.

2.2.3.3 Feature Extraction

After preprocessing, the information contained in the signal need to be extracted and selected, so feature extraction is performed.
So, in table 1, in terms of feature extraction, the use of Mel-frequency ceptrum coefficients (MFCCs) stands out, with 26 articles of the total of 87. The MFCC technique is utilized to extract significant characteristics from the audio signal. The implementation of MFCC involves a few steps. The first one is known as pre-emphasis, which involves applying a high-pass filter to compensate for the suppressed frequencies during the acquisition of the audio signal. After this, framing and windowing are carried out, where the signal is divided into frames and each of these pieces is subjected to the action of a window, reducing ripple and smoothing the signal, in Table 1, the most commonly used is the Hamming window. The third step is the use of a Discrete Fourier Transform (DFT) to represent the frames in the frequency domain, then non-linear representations of the signal are extracted using a Mel band-pass filter where a Discrete Cosine Transform (DCT) is applied to separate the relationships in the logarithmic spectral magnitudes or to identify certain

features of the audio signal. [109]
In the feature extraction domain, the representation of the speech in a spectrogram, or heat map, is wildly used too in Table 1, with 20 articles in a total of 87. A spectrogram [110] refers to a coloured 2D representation of the frequency in relation to time, or a 3D representation if the amplitude is also represented, where the intensity of the colour varies according to the density. Methods such as band-pass filters (e.g. Mel band-pass filters) or Fourier transforms (e.g. STFT) can be used to build them.

2.2.3.4 Classification and Validation

For classification purposes, the classifiers most used in the articles mentioned in Table 1 can be grouped into two categories, namely deep learning classifiers and machine learning classifiers.
Deep learning classifiers include neural networks, including convolutional neural networks (CNN), artificial neural networks (ANN), Residual Neural Network (ResNET), Multilayer perceptron (MLP), among others. The group of machine learning classifiers includes algorithms such as Support Vector Machine (SVM), Random Forest (RF), k-nearest neighbors (k-NN), Decision Tree (DT), Naïve Bayes (NB), Linear Discriminant Analysis (LDA), Linear Regression, ensemble learning techniques (AdaBoost, Gradient Boost, etc.), among others.
For the classification of the selected features, the division of acquired biomedical data into three sets is widely observed, that is, training, validation, and testing. With this separation of data, it is possible to improve the performance of the chosen classifier. There is also a combination of several deep and machine learning algorithms in the same study, for example [81], as well as a comparison between performances obtained by the proposed algorithm and algorithms usually found in studies in the area, e.g. [72], [77] and [85], in order to conclude whether the proposed methodology is really an asset to the scientific field.
In the field of validation, the most used technique, in Table 1, was cross-validation (CV), of which Leave-One-Out cross-validation (LOO CV), Leave-one-subject-out cross-validation (LOSO CV), Leave-One-Person-Out cross-validation (LOPO CV) and k-fold cross-validation (11-fold CV, 10-fold CV and 5-fold CV) stand out.
In terms of evaluating the performance of the proposed algorithms, formulas such as accuracy, precision, F1-Score, sensitivity, specificity, Mean Absolute Error and Area Under the Curve are commonly used.

2.3 Conclusive notes

This systematic review of the literature can be classified based on the ultimate goal of the studies included, indicating that it can be separated into sections that focus on the diag-

nosis and prognosis of PD.

In the case of the prognosis of PD, the studies generally focus on predicting UPDRS score and Hoehn & Yard scale (H&Y), and, from there, the severity of the disease is known. As can be seen in Table 1, there are 6 articles focused on the prognosis and not the diagnosis or detection of the disease ([44], [22], [80], [83], [89], [91]), although in [89] the prognosis and diagnosis of PD was performed, the main goal remains the prognosis of PD. In the case of the diagnosis of PD, which is the majority of the articles presented in Table 1, the studies can be divided into two focuses, feature extraction, and classification methods, i.e. part of the studies focused on studying the acquisition of the best features and methods of extracting them, while the other part focused on the choice of the classifier algorithm, in order to obtain the best performance.

On the other hand, this systematic literature review shows a greater use of isolated and sustained vowels, compared to, for example, monologues or reading texts, to make the diagnosis. It is also noticeable, even though the difference between performances was not very significant, that the use of classifiers based on deep learning techniques or in ensemble learning techniques performs better than the other classifiers presented, which can be seen in articles such as [51] and [74]. Although it is also very noticeable that the SVM and RF machine learning algorithms stand out in terms of usage and performance when it comes to predicting the disease.

In addition to the use of neural networks, this SLR also shows the use of other bioinspired algorithms to select features, for example in the articles [31], [57], [59] and [61], algorithms such as Modified Grey Wolf, Genetic Algorithm and Discrete Artificial Bee Colony, demonstrating their good performance in feature selection.

In conclusion, evaluating the performance of the algorithms presented, the vast majority have an accuracy greater than at least 70%, with a large percentage of articles having an accuracy greater than 90%, some with values of 100%, demonstrating the efficiency of applying deep and machine learning techniques to the diagnosis and prognosis of Parkinson's disease.

Chapter 3

Speech and voice assessment approach

In this chapter, the methodology for the proposed approach is presented and discussed.

3.1 Data

To conduct this research, the MDVR-KCL database [107] (available online at `https://zenodo.org/records/2867216`), already briefly mentioned in Section 2.2, as used.

The database, entitled Mobile Device Voice Recordings at King's College London, was registered at King's College London Hospital, Denmark Hill, Brixton, London, between September 26 and 29, 2017, and published in 2019.

It consists of 73 high quality audio recordings, with a sample rate of 44.1 kHz and a bit depth of 16 Bit, of people with and without Parkinson's disease, organized into a control group (HC) and a study group (PD), the study group presenting recordings of 3 levels of Parkinson's disease, level 2, 3 and 4, considering a scale of 4 possible levels, where level 1 is the most initial level of the disease, in which the symptoms are not noticeable and level 4 the most severe of all, in which the disease is quite noticeable in the patient.
The aforementioned recordings were acquired as if they were a phone call, and the care with which they were recorded was also mentioned, taking into account the conditions of the surrounding environment.

The dataset is also divided into two different scenarios:

- **Read Text** - each patient read one or both of the folowings texts [107]:
  - **North Wind and the Sun**
    "The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other. Then the North Wind blew as hard as he could, but the more he blew the more closely did the traveler fold his cloak around him; and at last the North Wind gave up the attempt. Then the Sun shone out warmly, and immediately the traveler took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two."

  - **BNC – Tech. Engin. Computer applications in geography snippet**
    "[...] This is because there is less scattering of blue light as the atmospheric path length and consequently the degree of scattering of the incoming radiation is reduced. For the same reason, the sun appears to be whiter and less orange-coloured as the observer's altitude increases; this is because a greater proportion of the sunlight comes directly to the observer's eye. Figure 5.7 is a schematic representation of the path of electromagnetic energy in the visible spectrum as it travels from the sun to the Earth and back again towards a sensor mounted on an orbiting satellite. The paths of waves representing energy prone to scattering (that is, the shorter wavelengths) as it travels from sun to Earth are shown. To the sensor it appears that all the energy has been reflected from point P on the ground whereas, in fact, it has not, because some has been scattered within the atmosphere and has never reached the ground at all. [...]"

- **Spontaneous Dialogue** - each patient speaks during a phone call spontaneously, answering questions about places of interest, local traffic or personal interests [107].

The 73 recordings are divided by both groups, HC and PD, and also by both scenarios, these divisions are illustrated in Table 1.

Table 1: Number of recordings per scenario

	Read Text	Spontaneous Dialogue
HC	21	21
PD	16	15

Among the Parkinson's disease groups, patients are also separated into the different levels of the disease mentioned above. This division is shown in Table 2.

Table 2: Number of level of PD per scenario

	Read Text	Spontaneous Dialogue
PD2	8	8
PD3	6	6
PD4	2	1

3.2 Preprocessing

According to the online page of the MDVR-KCL dataset, the database was acquired within the reverberation radius, with a reverberation time of approximately 500 ms, so it can be considered "clean" [107].
The Systematic Literature Review presented shows that in the studies in which the aforementioned database is used, the preprocessing phase is reduced, usually featuring only

segmentation or downsampling phases, and may even be absent.
So, taking both information into account, for this approach, no preprocessing steps were carried out on the data extracted from the dataset.

3.3 Feature Extraction

For the feature extraction phase, the Antropy [111] and the Disvoice [112] packages, both Python frameworks, are used.

Using the Antropy package [111], it is possible to extract features related to the time-domain. It is divided into two sets, the first with entropy functions and the second with fractal dimension analysis functions. From the entropy domain, the functions *Spectral entropy* (se), *Hjorth mobility* and *complexity*, *Number of zero-crossings* (nzc), *Permutation entropy* (PE) and *Singular value decomposition entropy* (SVDE) will be used and from the fractal dimension group, the *Petrosian fractal dimension* (pfd), *Katz fractal dimension* (kfd), *Higuchi fractal dimension* (HFD) and *Detrended fluctuation analysis* (DFA) functions will be applied. As an example of the use of this type of functions in the field of Parkinson's disease detection, we can mention some examples found in the literature review, such as the use of entropy functions, such as nzc, in [57], and fractal dimension functions, such as DFA, in [50] and [74].

The DisVoice package [112] is a framework specialized in the analysis and extraction of features, from audio files, whether they are sustained vowels or continuous speech. It can be used to extract features, through static and dynamic matrices, from the phonological and phonation fields, as well as from the articulation, glottal and prosody domains. A review of the literature shows that this type of feature has been used repeatedly to assess speech in Parkinson's patients. One example is [58], in which articulation, phonation and prosody features were used to distinguish between Parkinson's patients and healthy people, and glottal features were used to obtain a description of the information present in the voice. Other examples include [35], [51], [64] and [22].

In this study, using the DisVoice structure, phonation [113] [114], prosody [114] [115] and glottal features [116] will be used to analyze continuous speech, by extracting the features from the static matrix of each set. The phonation and prosody features include features from the frequency and amplitude domains. In the glottal features, features related to glottal cycles are extracted, in the sense that times such as opening quotient (OQ), normalized amplitude quotient (NAQ), the difference between the first two harmonics of the glottal flow signal (H1H2) and the harmonic richness factor (HRF), i.e., the ratio between the sum of the amplitude of the harmonics and the amplitude of the fundamental frequency, are evaluated.

The Antropy and Disvoice packages are described more briefly in Table 3.

Table 3: Type of features

Package	Feature Group	Description	nº
Antropy	Entropy	Time series domain	6
	Fractal dimension	Time series domain	4
Disvoice	Phonation	(7 descriptors) x (4 functionals: mean, std, skewness, kurtosis)	28
	Prosody	Duration, fundamental frequency and energy	103
	Glottal	Glottal source ((9 descriptors) x (4 functionals: mean, std, skewness, kurtosis))	36
		Total	177

During this stage, the audio files were analyzed individually and it was possible to extract a total of 168 features, considering the four groups of Antropy, Phonation, Prosody and Glottal features.

The number of features is lower than that mentioned in the table 3, because *kfd* feature, from the Antropy group, and *skw1Evoiced*, *kurtosis1Evoiced*, *skw1Eunvoiced*, *kurtosis1Eunvoiced*, *skwlastEunvoiced*, *kurtosislastEunvoiced*, *skwlastEvoiced* and *kurtosislastEvoiced* features, from the Prosody group, showed errors for some patients when they were extracted, so it was decided to eliminate them. The updated number of features under study can be seen in table 4 and the features in study can be seen in Appendix A.1.

Table 4: Type of features after Feature Extraction step

Package	Feature Group	nº
Antropy	Entropy	6
	Fractal dimension	3
Disvoice	Phonation	28
	Prosody	95
	Glottal	36
	Total	168

During the extraction phase of the features from the Glottal group in the DisVoice package, failures were observed for some patients. Given that all features in the group failed for a certain number of patients, it was decided to exclude the patient recordings in which these failures were observed, rather than excluding the features as in the previous case. Therefore, 4 recordings were eliminated from the HC group for the Spontaneous Dialogue scenario, 2 recordings from the PD group for the Read Text scenario, and 2 recordings from the same group for the Spontaneous Dialogue scenario.
The updated number of recordings under study is given in the Table 5.

Table 5: Number of recordings per scenario after Feature Extraction step

	ReadText	SpontaneousDialogue
HC	21	17
PD	14	13

3.4 Feature Selection

After the feature extraction phase for each patient, a feature selection phase will be carried out. In this phase, the features with the greatest weight will be selected to distinguish between the HC group and the PD group and also between the different levels of the disease.

To make this selection possible, two types of tests will be applied. The first test aims to analyze and understand the type of distribution of the data acquired for each feature, and the second test in order to understand and determine the relationship between the different groups and thus be able to select the most important features. This second test will differ between the diagnostic phase, i.e., the distinction between the HC and PD groups, and the prognostic phase, i.e., the distinction between the control group and the different levels of the disease present in the PD group.

The first test to be carried out will be the Kolmogorov-Smirnov test, in order to understand the type of data distribution. The second test are the Mann-Whitney U test, for the group HC and PD, and the Kruskal-Wallis H-test, for the different levels, in order to determine the relationship between the features in study. Basically, the difference between these two tests is the number of sets of data each is able to analyze. The Mann-Whitney U-test can only make a comparison between two sets, while the Kruskal-Wallis H-test can do it for two or more sets [117].

3.5 Classification

After extracting and selecting the features, a classification phase will be performed.
This classification phase will first be performed for the features selected in the previous phase and then a classification phase will be implemented with different groups of features, a total of 15 groups, which differ from the group with the most important features, in order to understand the impact and weight that each group will have in this last phase.

A total of 10 machine learning classifiers will be used to classify the features, including three variations of the Bagging model, as well as the Decision Tree (DT), Support vector machine (SVM-SVC), Adaptive Boosting (AB), Random Forest (RF), Gradient Boosting (GB), Logistic Regression (LR) and Gaussian Naive Bayes (GNB) models.

Bagging models and Boosting models are ensemble learning methods, and the major variation between them is the way in which they perform the training phase. The former uses data replacement techniques and the latter uses data sequentially [118]. For this approach, three variations of a bagging model will be used, as already mentioned, i.e., a variation of the model using the K-nearest neighbors algorithm and two more versions using the Decision Tree model, in which the difference between the two will be related to whether or not the bootstrap argument is activated, something that will change the way

the classifier uses the data for training. In order to make it easier to distinguish between these three variations of the model, they will be referred to as BG-KNN, BG(T)-DT and BG(F)-DT, respectively. With regard to Boosting models, two models will be used for this approach: the Adaptive Boosting model [119] and the Gradient Boosting model. The last two classifiers mentioned, Logistic Regression and Gaussian Naive Bayes, although they differ, they are models based on a probabilistic approach [120], the latter using a Gaussian distribution.

The hyperparameters used for each classifier are described in Table 6.

Table 6: Classifier's hyperparameters

Classifier	Hyperparameters
BG-KNN	Estimator: K-nearest neighbor Number of learners: 87
BG(F)-DT	Estimator: Decision Tree Number of learners: 87 Bootstrap: False
DT	Criterion: Gini Impurity Maximum leaf nodes: 20
SVM-SVC	Kernel function: RBF *C* parameter: 3
AB	Number of learners: 7 Learning rate: 1 Algorithm: SAMME
BG(T)-DT	Estimator: Decision Tree Number of learners: 10
RF	Number of learners: 100
GB	Number of learners: 100 Learning rate: 0.1 Maximum number of levels: 3
LR	Solver algorithm: Liblinear Maximum number of iterations=100
GNB	Default hyperparameters

The results obtained for this classification phase will be evaluated using evaluation metrics such as accuracy, precision and recall.

3.6 Conclusive notes

The approach proposed for the assessment of parkinsonian speech has been described in this chapter.

The next chapter, Chapter 4, describes and discusses the results obtained from applying this approach. It will be divided into phases, with a first phase presenting the feature extraction phase, a second and third phase where the results obtained for the diagnosis and prognosis of the disease are presented and discussed, and a final phase where some final conclusions are made.

Chapter 4

Results and Discussion

In this chapter, the results obtained during the research will be presented and discussed. The information will be presented according to the following breakdown:

- **Diagnosis** - Evaluation between the control group (HC) and the PD group;
- **Prognosis** - Evaluation between the control group (HC) and the different levels of the disease (PD scores).

In both cases, the data was evaluated for the two scenarios.

4.1 Diagnosis

4.1.1 Statistical analysis and feature selection

After extracting all features, the Kolmogorov-Smirnov test was performed. The purpose of this step was to study the distribution of the data. Using a significant level of 0.05 ($p\text{-}value \leq 0.05$), the null hypothesis for the Kolmogorov-Smirnov test was rejected for 113 features in the case of the HC group and 116 features in the case of the PD group for the Read Text scenario, and for 111 features in the HC group and 113 features in the PD group for the Spontaneous Dialogue scenario.

After checking the distribution of the data, the Mann-Whitney U test was also applied in order to understand the importance of each feature under study. Given the same significant level of 0.05, the null hypothesis was rejected for 62 features in the Read Text scenario and for 31 features in the Spontaneous Dialogue scenario.

The most important features for both scenarios can be observed in Table 1 and Table 2, respectively, and the result of Mann-Whitney U test can be consulted in Appendix A.2.

Table 1: Most important features after Mann-Whitney U test for Read Text

Features
mobility; complexity; SVDE; stddurvoiced; se; UVU; VVU; PU; UP; skwdurpause; kurtosisdurpause; global avg std H1H2; global std avg; H1H2; nzc; maxdurvoiced; avgdurvoiced; global avg avg H1H2; Fotiltmax; std DDFo; stdEunvoiced; avgmseEunvoiced; global skewness avg H1H2; std DFo; skewness DFo; mindurvoiced; Fotiltstd; avgdurpause; Fokurt; global avg std QOQ; global avg avg HRF; Vrate; stdtiltEvoiced; avgtiltEvoiced; Fomsemax; std ppq; Fostd; skwtiltEvoiced; kurtosisEunvoiced; Fomax; avgdurunvoiced; skwdurunvoiced; min1Evoiced; stdmseEunvoiced; avg logE; global std avg NAQ; avglastEunvoiced; HFD; global std std NAQ; Fomsemin; global avg std HRF; Foavg; avgEunvoiced; maxlastEunvoiced; global kurtosis avg H1H2; maxdurpause; global skewness std H1H2; global avg avg NAQ; global avg std NAQ; avg ppq; Fotiltskw; Fomseku; stdtiltEunvoiced.

Table 2: Most important features after Mann-Whitney U test for Spontaneous Dialogue

Features
se; complexity; global avg avg H1H2; mobility; SVDE; global avg std H1H2; global std avg H1H2; global skewness std NAQ; global skewness avg NAQ; avgEunvoiced; global skewness avg H1H2; Fostd; std ppq; global kurtosis std NAQ; global avg std HRF; avg logE; Fomsemax; avgtiltEvoiced; std DFo; global std avg HRF; global kurtosis avg NAQ; Fomsestd; skwtiltEvoiced; std DDFo; skwmseEvoiced; global kurtosis avg H1H2; Fomax; Fotiltmin; DAF; stdtiltEvoiced; global std std HRF.

Box plots were then made for each of the features considered to be the most important, i.e., features that were rejected by the null hypothesis in the Mann-Whitney U test. The purpose of these box plots is to better understand and visualize the data for each feature and the relationships between the HC and PD groups.

Figures 1 and 2 show the box plots of the 6 most important features for the Read Text and Spontaneous Dialogue scenarios, respectively, taking into account previous tests. The box plots of the remaining features can be found in the Appendix A.4.

In both scenarios, it can be seen that the group with Parkinson's disease has a lower value, for most of the features mentioned, than the control group. The difference of values in both groups is well illustrated, making it viable to distinguish between a healthy person and a patient with Parkinson's disease.

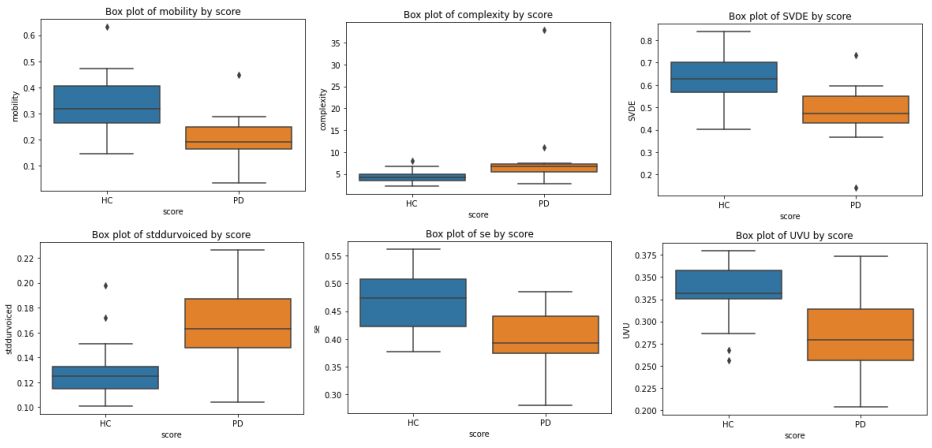

Figure 1: Box plots of the 6 main features of diagnostic phase for the Read Text

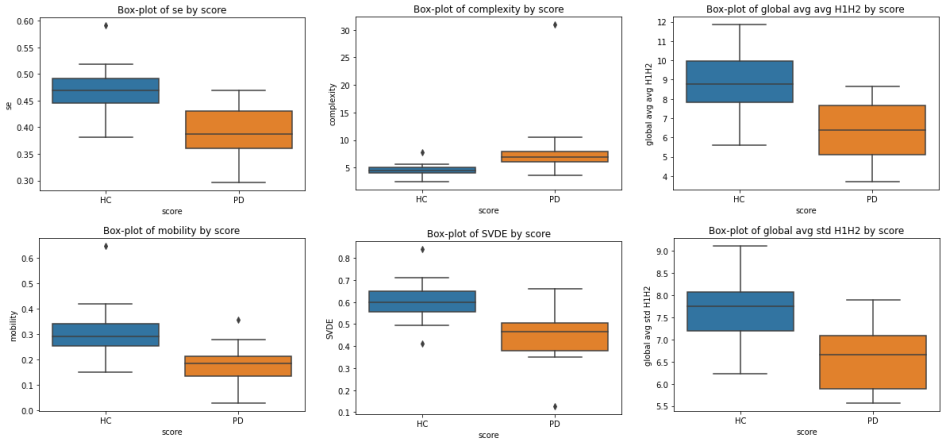

Figure 2: Box plots of the 6 main features of diagnostic phase for the Spontaneous Dialogue

4.1.2 Classification

After selection, the most significant features were classified using 10 different classification models, i.e., BG-KNN, BG(F)-DT, DT, SVM-SVC, AB, BG(T)-DT, RF, GB, LR and GNB.

The classification process was made between the control group (HC) and the Parkinson's group (PD), with 60% of the data for training and 40% for testing.

4.1.2.1 Main features

For this classification, 62 features were used in the Read Text scenario and 31 features for the Spontaneous Dialogue scenario. The performance results, for this first classification, of each of the classifiers are illustrated in Tables 3 and 4.

Examining Tables 3 and 4 it is clear to see the very good classification results obtained for both the Read Text and Spontaneous Dialogue scenarios, highlighting the excellent performance of the BG(F)-DT, DT, AB, BG(T)-DT, RF and GB models given the 100% values achieved for the three evaluation metrics. Also noteworthy is the performance of the remaining models, i.e., the BG-KNN, SVM-SVC, LR and GNB models, which, despite having lower accuracy values, always have precision values above 50%, something we can consider to be good. When we analyze these last four models, we can see the variation in both accuracy and precision values between the two scenarios under evaluation.

Comparing the first scenario, i.e., Read Text, with the second scenario, Spontaneous Dialogue, there is an improvement for the SVM-SVC and LR models and a worsening for the other two models, especially in terms of precision, which is extremely notable, for example, for SVM-SVC, where there is an increase in accuracy from 64% to 83% and in

precision from 67% for HC and 50% for PD to 78% and 100%, respectively.

Table 3: Performance results (%) of the most important features for Read Text

Classifier	Accuracy	Precision		Recall	
		HC	PD	HC	PD
BG-KNN	71	73	67	89	40
BG(F)-DT	100	100	100	100	100
DT	100	100	100	100	100
SVM-SVC	64	67	50	89	20
AB	100	100	100	100	100
BG(T)-DT	100	100	100	100	100
RF	100	100	100	100	100
GB	100	100	100	100	100
LR	86	89	80	89	80
GNB	86	89	80	89	80

Table 4: Performance results (%) of the most important features for Spontaneous Dialogue

Classifier	Accuracy	Precision		Recall	
		HC	PD	HC	PD
BG-KNN	67	67	67	86	40
BG(F)-DT	100	100	100	100	100
DT	100	100	100	100	100
SVM-SVC	83	78	100	100	60
AB	100	100	100	100	100
BG(T)-DT	100	100	100	100	100
RF	100	100	100	100	100
GB	100	100	100	100	100
LR	100	100	100	100	100
GNB	83	86	80	86	80

4.1.2.2 Feature comparison study

Following this first classification phase, a comparison study was conducted between the different feature groupings used when extracting features, i.e., between the *Antropy* (A), *Phonation* (Ph), *Prosody* (Pr) and *Glottal* (G) groups.
For this study, classification was performed for each group and possible combinations between them. The performance obtained for this classification phase is illustrated in Appendix A.5, Table 9.

By analyzing the data presented in Table 9, for the diagnosis phase, i.e., that includes an assessment of the control group and the group with Parkinson's disease, for both scenarios, Read Text and Spontaneous Dialogue, a good classification was obtained, with the classifiers used generally being able to distinguish between Parkinson's cases and control cases.

Of the machine learning algorithms used for this stage of classification, five stand out for obtaining accuracy values of 100% for all the groups of features tested, but on the

other hand, despite obtaining accuracy values of over 50%, the SVM-SVC classifier for the Spontaneous Dialogue scenario stands out in the negative, as it was unable to distinguish between Parkinson's and control cases for 10 of the 15 groups of features tested.

Of the groups of features tested, it can be seen that the *Antropy*, *Phonation*, *Prosody*, and *Glottal* sets obtain good accuracy values when used in isolation, with the last three sets standing out in relation to the first set, as they have higher values than the latter for the majority of classifiers. Taking the remaining groups into account, i.e., where the aforementioned sets are grouped together, globally, although this is not observed for all the classifiers and the values obtained do not show a very significant difference, it is noticeable that for the groups where the *Antropy* set is hidden, a slightly higher accuracy value is obtained compared to the groups where it is included, an example of a classifier in which this difference is notable is the GNB algorithm, for the *Total* group, i.e., the group in which the *Antropy+Phonation+Prosody+Glottal* sets are included, this algorithm obtained an accuracy of 86% and 83% for the Read Text and Spontaneous Dialogue scenarios respectively, and when we remove the *Antropy* set from this group, values of 93% and 92% were obtained respectively, something that proves the point made above, the same is also observed between the *Phonation+Glottal* group and the *Antropy+Phonation+Glottal* group, where the former obtained an accuracy of 100% and 83% for the two scenarios respectively and the latter obtained values of 86% and 75%, respectively.

This variation in values is illustrated in figures 3 and 4, and in addition to the example above, the behavior of two other classifiers, BG-KNN and SVM-SVC, is also displayed for the same feature sets.

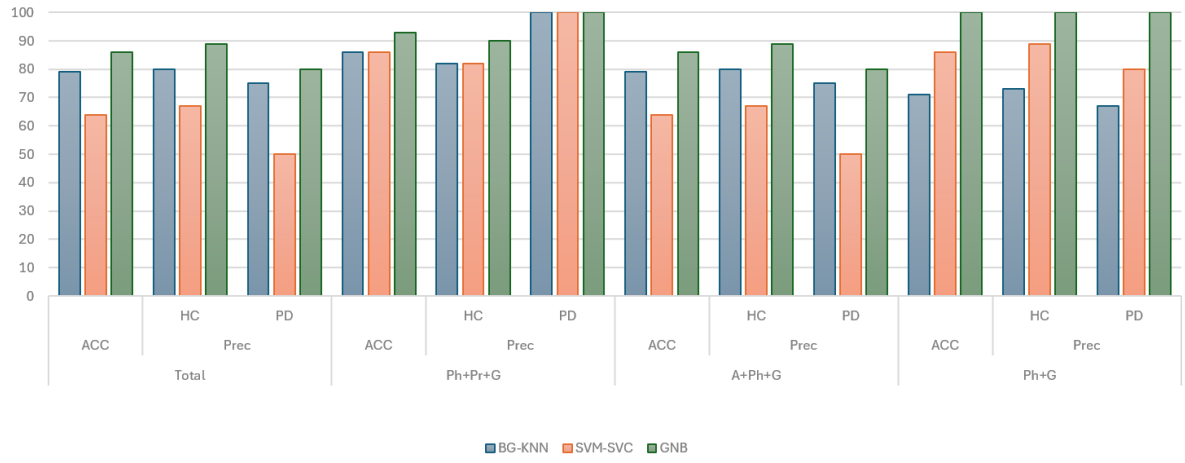

Figure 3: Performance results variations for Diagnosis - Read Text

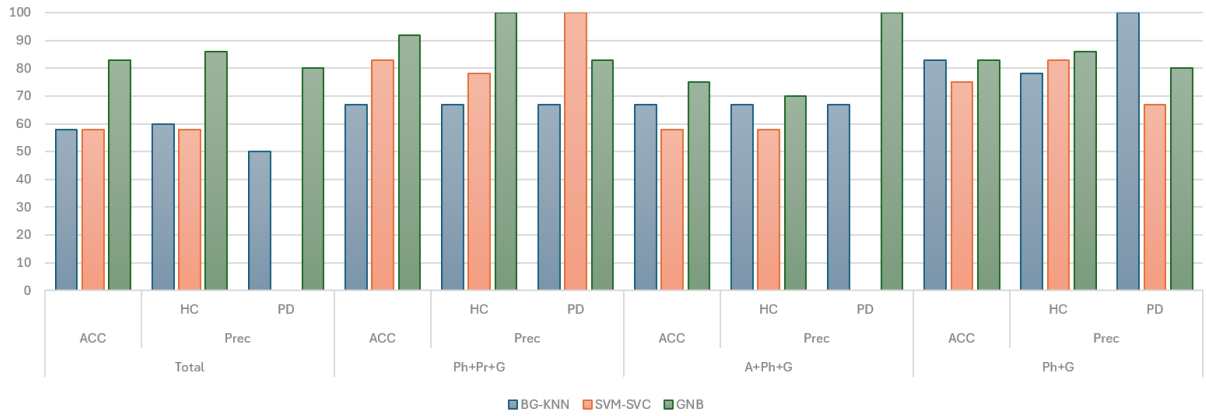

Figure 4: Performance results variations for Diagnosis - Spontaneous Dialogue

4.2 Prognosis

4.2.1 Statistical analysis and feature selection

After extracting all the features, as in the previous phase, the Kolmogorov-Smirnov test was conducted to assess the distribution of the data. Using a significant level of 0.05 (p-*value* $\leq$ 0.05), the null hypothesis for the Read Text scenario was rejected for 109, 88 and 50 features at the PD2, PD3 and PD4 levels, respectively, while for the Spontaneous Dialogue scenario it was rejected for 105, 81 and 43 features at the PD2, PD3 and PD4 levels, respectively. After checking the distribution of the data, the Kruskal-Wallis H-test was also carried out in order to understand the importance of each feature under study. Given the same significant level of 0.05, the null hypothesis was rejected for 15 features in the Read Text scenario and for 8 features in the Spontaneous Dialogue scenario.

The most important features for both scenarios can be observed in Table 5 and Table 6, respectively, and the results of Kruskal-Wallis H-test can be consulted in Appendix A.3.

Table 5: Most important features after Kruskal-Wallis H-test for Read Text

Features
avgdurpause; PU; UP; PVU; Vrate; mobility; SVDE; kurtosisdurunvoiced; complexity; VP; se; minlastEunvoiced; stddurpause; stdlastEunvoiced; avgtiltEvoiced.

Table 6: Most important features after Kruskal-Wallis H-test for Spontaneous Dialogue

Features
mobility; complexity; SVDE; Vrate; Fotiltmin; HFD; se; global skewness avg QOQ.

Box plots were then made for each of the features considered most important after the Kruskal-Wallis H-test. The purpose of these box plots is to better understand and visualize

the data for each feature and the relationships between the control group and the various Parkinson's disease scores.

Figures 5 and 6 show the box-plots for the 6 most important features for the Read Text and Spontaneous Dialogue scenarios, respectively, taking into account the tests carried out previously. The box-plots of the remaining features can be found in Appendix A.4.

In this case, taking into account the figures, it is possible to see that there is a clear distinction between an initial level of the disease and a more advanced level, as can be seen in the values presented for groups PD3 and PD4, more severe levels of PD, and for group PD2, a level where the disease is at an early stage and the symptoms are not so noticeable.

Having said this, it is also possible to get a brief idea of why there are cases of false diagnosis for patients in the early stages of Parkinson's disease, when comparing the values for the HC group with those for the PD2 group, it can be seen that there is no significant variation, which can make it difficult to determine whether the diagnosis is true or not.

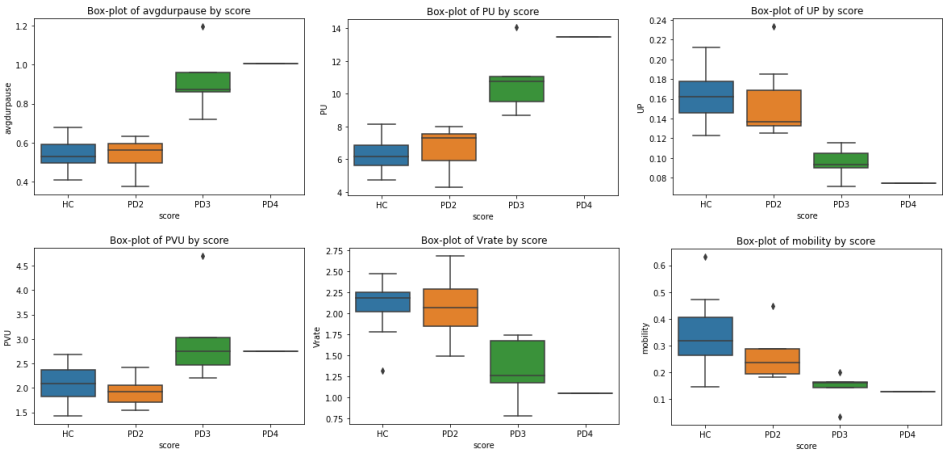

Figure 5: Box plots of the 6 main features of prognosis phase for the Read Text

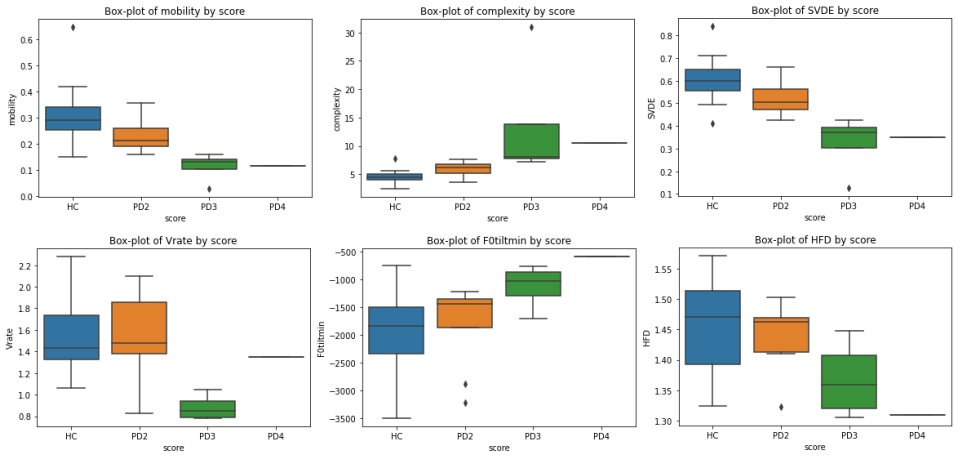

Figure 6: Box plots of the 6 main features of prognosis phase for the Spontaneous Dialogue

4.2.2 Classification

After selection, the most significant features were classified using 10 different classification models, i.e., BG-KNN, BG(F)-DT, DT, SVM-SVC, AB, BG(T)-DT, RF, GB, LR and GNB.

The classification process was made between the control group (HC) and the different levels available in the Parkinson's group (PD), with 60% of the data for training and 40% for testing.

Since there was only one element with a parkinsonism level of 4, i.e., the most severe level, and, in order to be able to be included in the classification phase, this patient was considered to be a level 3 patient, so a classification was made between a level 2 and a level equal to or greater than 3 ($\geq$ PD3).

4.2.2.1 Main features

For this classification, 15 features were used in the Read Text scenario, with this value being 8 for the Spontaneous Dialogue scenario. The performance results, for this first classification, of each of the classifiers are illustrated in Tables 7 and 8.

Analyzing the values shown in Table 7 and Table 8, it can be seen that, in general, very good results were obtained for this classification phase, with the repeated accuracy and precision values of 100% for at least 5 of the 10 algorithms applied standing out in a very positive way. However, although in general terms the performance values obtained can be considered a success, it is also necessary to highlight the failures that occurred with the BG-KNN, SVM-SVC and LR models, especially when it came to distinguishing between the two levels of the disease, because, despite having high accuracy values, it is notable that they failed to classify at least one of the levels, given the precision values of 0.

A brief comparison between the values obtained for both scenarios, i.e., the first scenario, Read Text, and the second, Spontaneous Dialogue, shows that for the BG-KNN, SVM-SVC, AB, LR and GNB models, there is a drop in the values obtained for the evaluation metrics, as an example of which is the BG-KNN model, which obtains, for first scenario, precision values of 80%, 100% and 50% for the HC, PD2 and $\geq$ PD3, respectively, and 62%, 67% and 0%, respectively, for the second scenario, a significant worsening, especially for the distinction between PD cases, where it is no longer performed for the last level and shows a significant decline for the other level.

Table 7: Performance results (%) of the most important features for ReadText

Classifier	Accuracy	Precision			Recall		
		HC	PD2	≥ PD3	HC	PD2	≥ PD3
BG-KNN	79	80	100	50	89	67	50
BG(F)-DT	100	100	100	100	100	100	100
DT	100	100	100	100	100	100	100
SVM-SVC	64	67	0	50	89	0	50
AB	93	90	100	100	100	67	100
BG(T)-DT	100	100	100	100	100	100	100
RF	100	100	100	100	100	100	100
GB	100	100	100	100	100	100	100
LR	79	75	0	100	100	0	100
GNB	86	89	67	100	89	67	100

Table 8: Performance results (%) of the most important features for Spontaneous Dialogue

Classifier	Accuracy	Precision			Recall		
		HC	PD2	≥ PD3	HC	PD2	≥ PD3
BG-KNN	58	62	67	0	71	67	0
BG(F)-DT	100	100	100	100	100	100	100
DT	100	100	100	100	100	100	100
SVM-SVC	58	58	0	0	100	0	0
AB	83	78	100	100	100	33	100
BG(T)-DT	100	100	100	100	100	100	100
RF	100	100	100	100	100	100	100
GB	100	100	100	100	100	100	100
LR	67	67	0	100	86	0	100
GNB	75	83	50	100	71	67	100

4.2.2.2 Feature comparison study

After this first classification phase, a comparison study was carried out between the different feature groupings used when extracting features. For this study, classification was carried out for each group and possible combinations between them. The performance obtained is illustrated in Appendix A.5, Table 10.

By analyzing the data presented in the Table 10, for the prognosis phase, i.e., the evaluation between the different levels of the disease and the control group, in both scenarios, very satisfactory accuracy values were also obtained when distinguishing between the levels of Parkinson's presented for evaluation, with 4 of the 10 algorithms used standing out positively for their accuracy values of 100% for the different groupings of features, but on the other hand, for this prognostic phase, it is also noteworthy that some algorithms failed to classify and distinguish between the different levels of the disease, for example, the SVM-SVC and BG-KNN algorithms were unable to perform a good classification for the levels of Parkinson's presented for most of the feature groups, and this failure was also observed in the AB and LR algorithms for at least 2 of the 15 feature groups subject to classification.

With regard to the feature sets and their respective groupings, it is possible to observe the same type of behaviour mentioned during the diagnosis phase of the disease, i.e., better accuracy values are obtained when the *Antropy* set is not included, as we can observe in the Read Text scenario, although for this prognosis phase it is possible to observe the opposite behaviour, even if they are in a few isolated cases, i.e., it is possible to observe that for some cases, the inclusion of the Antropy set was a plus, observing the classification obtained for the Spontaneous Dialogue scenario, has as an example the GNB classifier, for the *Total* group obtained a precision of 100%, already for the group without the *Antropy* obtained a value of 92%, a descent from the precision of the algorithm, we also have the BG(T)-DT classifier for the *Antropy+Phonation+Prosody* group and the *Phonation+Prosody* group, where the values were 100% and 92%, respectively. For the behavior in which the omission of the *Anthropy* group is an advantage, for the same scenario, we have the example of AB and of SVM-SVC for the *Total* group and for the *Phonation+Prosody+Glotal* group, where there is an increase in the accuracy value, from 83% to 92% and 58% to 67%, respectively.

Figures 7 and 8 show the variations in results due to the absence or not of the *Anthropy* group. For the Read Text scenario, the behaviors of the BG-KNN, SVM-SVC and GNB classifiers are illustrated for the *Total*, *Ph+Pr+G*, *A+Ph+Pr* and *Ph+Pr* feature groups, for the Spontaneous Dialogue scenario, the SVM-SVC, AB and GNB classifiers are shown for the *Total* and *Ph+Pr+G* feature sets.

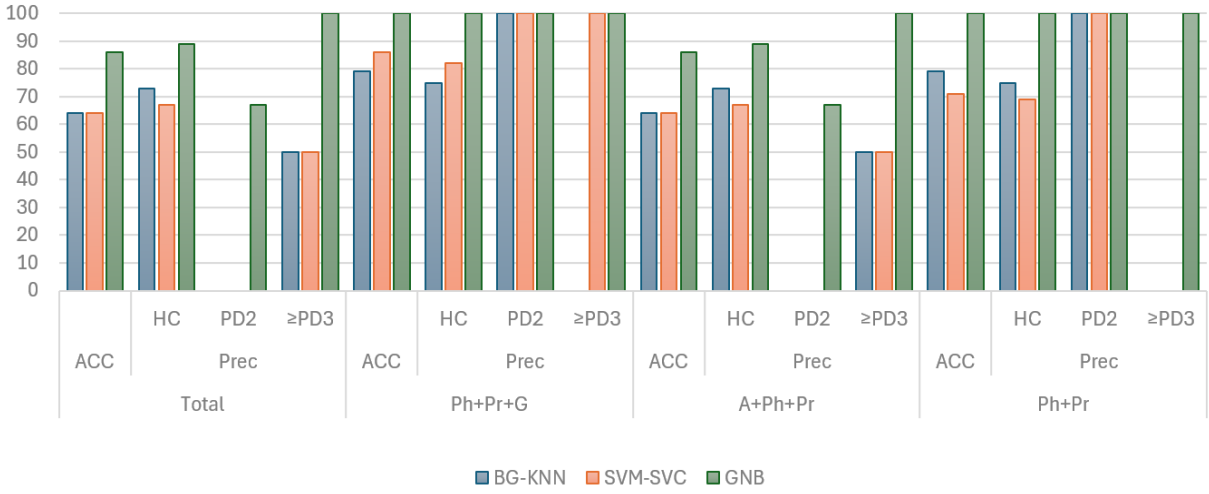

Figure 7: Performance results variations for Prognosis - Read Text

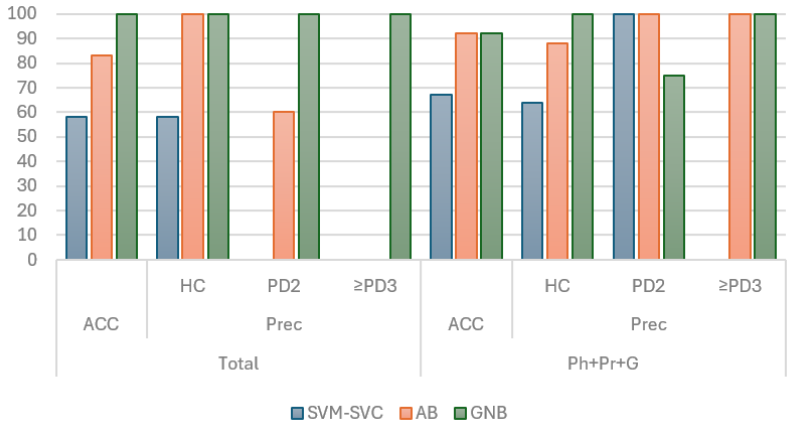

Figure 8: Performance results variations for Prognosis - Spontaneous Dialogue

4.3 Conclusive notes

In conclusion, the results show that it is feasible to detect and prognosticate Parkinson's disease by evaluating a patient's continuous speech.

In general, it was possible to achieve a good classification for both scenarios and for both phases presented for study. The Read Text scenario stands out, as the best performance values were obtained for both phases, although a much very satisfactory classification was also obtained for the Spontaneous Dialogue scenario.

Regarding the comparative study carried out between different groups of features, it can be concluded that, although the *Antropy* set has obtained excellent classification values and can be an added value in some cases, the other three sets stood out positively in relation to it, even presenting better results when the latter is not included in the classification.

With regard to the classifiers used, the DT, BG(F)-DT, AB, RF, GB classifiers stand out positively in this classification phase, since they obtained accuracy values of 100% for most of the feature groups used and for both scenarios and the diagnosis and prognosis phases carried out. On the other hand, the BG-KNN and SVM-SVC classifiers stand out negatively, since they were the classifiers that showed the most errors when classifying cases of Parkinson's disease and also between the different levels of the disease.

As already mentioned throughout this chapter, the high values obtained for the metric evaluation conducted in the classification phase are extremely noteworthy, since various values of 100% were achieved, whether for accuracy or precision. This type of value could create some doubt or question regarding the evaluation of the results achieved for this approach, but if we take a brief look at the Systematic Literature Review, it is possible to observe, with some regularity, performance values within the same range as those obtained, either for the same type of classifiers used or for approaches in which the same

database was used, for example the articles [88] and [89]. Both use the same dataset, MDVR-KCL. In [88], the disease is diagnosed using a Deep convolution neural network (DCNN) model for the selection and classification of spectrum images of voice samples, where accuracy values of 99.48% were obtained for the training phase. In [89], a more similar approach to the one presented in terms of classification, both the diagnosis and prognosis of the disease are made with accuracy values of 98.62% and precision between 80% and 100%. In this case, a voice activity detection (VAD) technique is applied as a pre-processing step, where voice segments and pause segments are identified, then a wavelet scattering (WS) transform is applied to extract robust and discriminative features of these segments, which are then classified by seven machine learning algorithms, whose optimal hyperparameters were chosen using a Bayesian optimization (BO) approach.
By analyzing these two articles, we can see a difference with the proposed approach, in addition to the difference between applications and methods for obtaining features and subsequent classification, which is the signal evaluated, in which, in both approaches, segments of the signal are evaluated and not the signal as a whole, something that occurs in this proposed approach.

The analysis of this different performances obtained also serves to reinforce the importance of using different evaluation metrics, since a high accuracy value does not always mean a good prediction by the classifier, something that is quite noticeable in the various classification stages presented in this chapter.

Chapter 5

Conclusions and Future Work

This final chapter will elucidate the final conclusions drawn from the research work described in the previous chapters, as well as possible questions and future work with a view to improving the approach presented.

5.1 Conclusion

This dissertation proposes an approach for Parkinson's disease, diagnosing and prognosticating the disease by assessing voice and speech features using machine learning techniques. This type of approach offers added value, both in terms of time and in terms of a possible reduction in false diagnoses, as well as providing an insight into the patient's level of illness, something that can complement the important analysis made by a health professional.
In order to conduct this research, objectives were defined taking into account the literature found in the area:

- Research and analysis in the literature review on possible open questions in the field.
- The formulation of an approach to the diagnosis and prognosis of Parkinson's disease through the evaluation of speech and voice.
- Identification of features with the most impact on the diagnosis and prognosis of Parkinson's disease through voice analysis.
- Compare performance values for different groups of features, reducing computational time, without negatively influencing the final results.

They are summarized below.

The first objective is achieved in Section 2.2 of Chapter 2. It presents the Systematic Literature Review of the proposed topic, which included 87 studies on the detection and prognosis of Parkinson's disease through voice or speech assessment. Of these studies, the vast majority, around 81, present only a proposal for the diagnosis or detection of this neurological disease, and only 6 studies were found that performed prognosis, either in isolation or in conjunction with diagnosis. The SLR shows that some of the approaches presented mainly assess recordings of sustained vowels or small phonemes, and less recordings of continuous speech.

The second objective can be found in chapters 3 and 4. Chapter 3 presents and describes the method to be implemented. It proposes an approach to diagnosis and prognosis by analyzing and evaluating voice and speech using entropy, fractal dimension, phonation, prosody and glottal features. Chapter 4 presents and discusses the results obtained for this proposal. The conclusion is that the proposed approach is feasible and capable of performing both diagnosis and prognosis for Parkinson's disease.

The third objective can be found in Chapter 4. To achieve this, tests such as the Mann-Whitney U-test and the Kruskal-Wallis H-test were used, in order to assess the importance of each feature under study. With the completion of this objective, it was concluded that for each scenario and application, whether diagnostic or prognostic, there is a variation in the features chosen as most important by the tests, both in number and type.

The fourth objective is presented in Chapter 4. For this purpose, a comparison study is conducted for the different types of features (Anthropy, Phonation, Prosody and Glottal) and possible combinations between them, in order to establish their weight and influence on the results obtained. This study concluded that the use of groups from the DisVoice framework is an asset to the classification process. It was also concluded that the influence of the Antropy group is somewhat variable and depends on the application, be it diagnosis or prognosis, and the classifier in use, and can be an asset for some and a disadvantage for others.

5.2 Future Work

With the finalization of this project, additional questions and future possibilities were raised.

- Will the approach be reproducible for the assessment of other languages besides English?
- Despite not being considered as a variable, could the gender of the subjects being evaluated have had any influence on the results achieved?
- The realization of this approach to the prognosis of Parkinson's disease using a larger number of patients for the four levels of the disease.
- Realization of this approach using a combination of the features of static and dynamic matrices of the DisVoice framework.
- What changes could occur in the results of this approach if a preprocessing step was applied.
- The possibility of implementing this approach in the future, for example in a mobile application or a wearable device, in order to complement the clinical procedure already practiced.

Bibliography

[1] N. I. of Neurological Disorders and S. [online] www.ninds.nih.gov, "National institute of neurological disorders and stroke (2023). parkinson's disease," 2023 (Accessed on : 09 January 2024). [Online]. Available: https://www.ninds.nih.gov/health-information/disorders/parkinsons-disease 1, 6

[2] WHO, "Parkinson disease," 2023 (Accessed on : 09 January 2024). [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/parkinson-disease# 1, 6

[3] B. R. Bloem, M. S. Okun, and C. Klein, "Parkinson's disease," *The Lancet*, vol. 397, no. 10291, pp. 2284–2303, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S014067362100218X 1, 6

[4] Healthline, "Parkinson's mri: Diagnosing early onset through brain imaging," 2022 (Accessed on: 10 January 2024). [Online]. Available: https://www.healthline.com/health/parkinsons/parkinsons-mri 2

[5] N. I. of Neurological Disorders and Stroke, "Brain basics: Know your brain | national institute of neurological disorders and stroke," www.ninds.nih.gov, 2023 (Accessed on : 26 February 2024). [Online]. Available: https://www.ninds.nih.gov/health-information/public-education/brain-basics/brain-basics-know-your-brain 5

[6] W. G. Webb, "Organization of the nervous system i," *Neurology for the Speech-Language Pathologist*, pp. 13–43, 2017. 5

[7] C. Goetz, "The history of parkinson's disease: Early clinical descriptions and neurological therapies. cold spring harbor perspectives in medicine, [online] 1(1)." 2011. 6

[8] Z. Z. Sonkaya, M. Ceylan, and A. R. Sonkaya, "Speech characteristics of parkinson disease," *Medical Science and Discovery*, vol. 8, no. 12, p. 666–670, Dec. 2021. [Online]. Available: https://medscidiscovery.com/index.php/msd/article/view/645 6, 7

[9] P. Mounika and S. G. Rao, "Machine learning and deep learning models for diagnosis of parkinson's disease: A performance analysis," *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 11 2021. 7, 8, 12

[10] M. B. Akçay and K. Oğuz, "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers," *Speech Communication*, vol. 116, p. 56–76, 01 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167639319302262 8

[11] B. Ghojogh, M. N. Samad, S. A. Mashhadi, T. Kapoor, W. Ali, F. Karray, and M. Crowley, "Feature selection and feature extraction in pattern analysis: A literature review," *ArXiv (Cornell University)*, vol. abs/1905.02845, 2019. 8, 9

[12] Z. M. Hira and D. F. Gillies, "A review of feature selection and feature extraction methods applied on microarray data," *Advances in Bioinformatics*, vol. 2015, pp. 1–13, 2015. 9

[13] G. Sharma, K. Umapathy, and S. Krishnan, "Trends in audio signal feature extraction methods," *Applied Acoustics*, vol. 158, p. 107020, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X19308795 9

[14] F. Alías, J. C. Socoró, and X. Sevillano, "A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds," *Applied Sciences*, vol. 6, p. 143, 05 2016. [Online]. Available: https://www.mdpi.com/2076-3417/6/5/143 9

[15] M. A. Myszczynska, P. N. Ojamies, A. M. B. Lacoste, D. Neil, A. Saffari, R. Mead, G. M. Hautbergue, J. D. Holbrook, and L. Ferraiuolo, "Applications of machine learning to diagnosis and treatment of neurodegenerative diseases," *Nature Reviews Neurology*, vol. 16, pp. 440–456, 07 2020. 9, 10

[16] H. Gunduz, "Deep learning-based parkinson's disease classification using vocal feature sets," *IEEE Access*, vol. 7, pp. 115 540–115 551, 2019. 10, 15

[17] R. Sinha, N. Kaur, S. Gupta, and P. Thakur, "Diagnosis of parkinson's disease using hybrid ensemble technique," *2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE)*, pp. 1–5, 2023. 10, 36

[18] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001. 10

[19] S. Sharanyaa, P. N. Renjith, and K. Ramesh, "Classification of parkinson's disease using speech attributes with parametric and nonparametric machine learning techniques," *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, pp. 437–442, 2020. 10, 18

[20] T. Shah, "About train, validation and test sets in machine learning," Towards Data Science, 2017 (Last accessed on: 11 June 2024). [Online]. Available: https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7 10

[21] H. M and S. M.N, "A review on evaluation metrics for data classification evaluations," *International Journal of Data Mining Knowledge Management Process*, vol. 5, pp. 01–11, 03 2015. [Online]. Available: https://pdfs.semanticscholar.org/6174/3124c2a4b4e550731ac39508c7d18e520979.pdf 11, 12

[22] D. Hemmerling and M. Wojcik-Pedziwiatr, "Prediction and estimation of parkinson's disease severity based on voice signal," *Journal of Voice*, vol. 36, no. 3, pp. 439.e9–439.e20, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0892199720302319 12, 30, 42, 45

[23] L. Ali, C. Zhu, M. Zhou, and Y. Liu, "Early diagnosis of parkinson's disease from multiple voice recordings by simultaneous sample and feature selection," *Expert Systems with Applications*, vol. 137, pp. 22–28, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095741741930452X 15

[24] K. Polat, "A hybrid approach to parkinson disease classification using speech signal: The combination of smote and random forests," *2019 Scientific Meeting on Electrical-Electronics Biomedical Engineering and Computer Science (EBBT)*, pp. 1–3, 2019. 15

[25] C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, M. Tutuncu, T. Aydin, M. E. Isenkul, and H. Apaydin, "A comparative analysis of speech signal processing algorithms for parkinson's disease classification and the use of the tunable q-factor wavelet transform," *Applied Soft Computing*, vol. 74, pp. 255–263, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494618305799 15

[26] S. Bhati, L. M. Velazquez, J. Villalba, and N. Dehak, "Lstm siamese network for parkinson's disease detection from speech," *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1–5, 2019. 16

[27] Shivangi, A. Johri, and A. Tripathi, "Parkinson disease detection using deep neural networks," *2019 Twelfth International Conference on Contemporary Computing (IC3)*, pp. 1–4, 2019. 16

[28] J. C. Vásquez-Correa, T. Arias-Vergara, J. R. Orozco-Arroyave, B. Eskofier, J. Klucken, and E. Nöth, "Multimodal assessment of parkinson's disease: A deep learning approach," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1618–1630, 2019. 16

[29] M. Wodzinski, A. Skalski, D. Hemmerling, J. R. Orozco-Arroyave, and E. Nöth, "Deep learning approach to parkinson's disease detection using voice recordings and convolutional neural network dedicated to image classification," *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 717–720, 2019. 16

[30] D. Braga, A. M. Madureira, L. Coelho, and R. Ajith, "Automatic detection of parkinson's disease based on acoustic analysis of speech," *Engineering Applications of Artificial Intelligence*, vol. 77, pp. 148–158, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197618302045 17

[31] P. Sharma, S. Sundaram, M. Sharma, , A. Sharma, and D. Gupta, "Diagnosis of parkinson's disease using modified grey wolf optimization," *Cognitive Systems Research*, vol. 54, pp. 100–115, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389041718308726 17, 42

[32] B. Karan, S. S. Sahu, and K. Mahto, "Parkinson disease prediction using intrinsic mode function based features from speech signal," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 249–264, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0208521618305564 17

[33] O. Asmae, R. Abdelhadi, C. Bouchaib, S. Sara, and K. Tajeddine, "Parkinson's disease identification using knn and ann algorithms based on voice disorder," *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, pp. 1–6, 2020. 17

[34] L. Zahid, M. Maqsood, M. Y. Durrani, M. Bakhtyar, J. Baber, H. Jamal, I. Mehmood, and O. Y. Song, "A spectrogram-based deep feature assisted computer-aided diagnostic system for parkinson's disease," *IEEE Access*, vol. 8, pp. 35 482–35 495, 2020. 18

[35] T. V. Priya, S. Sivapatham, and A. Kar, "Parkinson's disease detection using multiple speech signals," *2020 IEEE 4th Conference on Information Communication Technology (CICT)*, pp. 1–5, 2020. 18, 45

[36] E. Majda-Zdancewicz, A. Dobrowolski, A. Potulska-Chromik, J. Jakubowski, J. Chmielińska, K. Białek, M. Nojszewska, and A. Kostera-Pruszczyk, "The use of non-linear acoustic analysis to objectively evaluate the voice of people with parkinson's disease," *2020 IEEE 21st International Conference on Computational Problems of Electrical Engineering (CPEE)*, pp. 1–4, 2020. 18

[37] S. Dávid, O. Anett, and I. Valálik, "Articulation correctness measurement of parkinson's disease using low resource-intensitve segmentation methods," *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pp. 000 121–000 124, 2020. 19

[38] B. Karan, S. S. Sahu, and K. Mahto, "Stacked auto-encoder based time- frequency features of speech signal for parkinson disease prediction," *2020 International Conference on Artificial Intelligence and Signal Processing (AISP)*, pp. 1–4, 2020. 19

[39] Z. Karapinar Senturk, "Early diagnosis of parkinson's disease using machine learning algorithms," *Medical Hypotheses*, vol. 138, p. 109603, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306987719314148 19

[40] A. H. Poorjam, M. S. Kavalekalam, L. Shi, J. P. Raykov, J. R. Jensen, M. A. Little, and M. G. Christensen, "Automatic quality control and enhancement for

voice-based remote parkinson's disease detection," *Speech Communication*, vol. 127, pp. 1–16, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167639320303113 19

[41] F. Cordella, A. Paffi, and A. Pallotti, "Classification-based screening of parkinson's disease patients through voice signal," *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–6, 2021. 20

[42] J. Ma, Y. Zhang, Y. Li, L. Zhou, L. Qin, Y. Zeng, P. Wang, and Y. Lei, "Deep dual-side learning ensemble model for parkinson speech recognition," *Biomedical Signal Processing and Control*, vol. 69, p. 102849, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809421004468 20

[43] M. Mohaghegh and J. Gascon, "Identifying parkinson's disease using multimodal approach and deep learning," *2021 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications (CITISIA)*, pp. 1–6, 2021. 20

[44] L. Naranjo, C. J. Pérez, Y. Campos-Roca, and a, "Monitoring parkinson's disease progression based on recorded speech with missing ordinal responses and replicated covariates," *Computers in Biology and Medicine*, vol. 134, p. 104503, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482521002973 20, 42

[45] F. Huang, H. Xu, T. Shen, and L. Jin, "Recognition of parkinson's disease based on residual neural network and voice diagnosis," *2021 IEEE 5th Information Technology,Networking,Electronic and Automation Control Conference (ITNEC)*, vol. 5, pp. 381–386, 2021. 21

[46] D. A. Vigneswari and J. Aravinth, "Parkinson's disease diagnosis using voice signals by machine learning approach," *2021 International Conference on Recent Trends on Electronics, Information, Communication Technology (RTEICT)*, pp. 869–872, 2021. 21

[47] M. Younis Thanoun and M. T. YASEEN, "A comparative study of parkinson disease diagnosis in machine learning - proceedings of the 4th international conference on advances in artificial intelligence," p. 23–28, 2021. [Online]. Available: https://doi.org/10.1145/3441417.3441425 21

[48] A. Ouhmida, O. Terrada, A. Raihani, B. Cherradi, and S. Hamida, "Voice-based deep learning medical diagnosis system for parkinson's disease prediction," *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, pp. 1–5, 2021. 21

[49] R. Fayad, M. Hajj-Hassan, G. Costantini, Z. Zarazadeh, V. Errico, A. Pisani, G. D. Lazzaro, M. Ricci, and G. Saggio, "Vocal test analysis for assessing parkinson's dis-

ease at early stage," *2021 Sixth International Conference on Advances in Biomedical Engineering (ICABME)*, pp. 171–174, 2021. 22

[50] R. Laila, L. Salwa, and R. Mohammed, "Detection of voice impairment for parkinson's disease using machine learning tools," *2020 10th International Symposium on Signal, Image, Video and Communications (ISIVC)*, pp. 1–6, 2021. 22, 45

[51] C. Quan, K. Ren, and Z. Luo, "A deep learning based method for parkinson's disease detection using dynamic features of speech," *IEEE Access*, vol. 9, pp. 10 239–10 252, 2021. 22, 42, 45

[52] B. Karan, S. S. Sahu, J. R. Orozco-Arroyave, and K. Mahto, "Non-negative matrix factorization-based time-frequency feature extraction of voice signal for parkinson's disease prediction," *Computer Speech Language*, vol. 69, p. 101216, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0885230821000231 23

[53] J. Goyal, Kh, P. nor, and T. C. Aseri, "A hybrid approach for parkinson's disease diagnosis with resonance and time-frequency based features from speech signals," *Expert Systems with Applications*, vol. 182, p. 115283, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421007144 23

[54] O. Karaman, H. Çakın, A. Alhudhaif, and K. Polat, "Robust automated parkinson disease detection based on voice signals with transfer learning," *Expert Systems with Applications*, vol. 178, p. 115013, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421004541 23

[55] Y. C. Tai, P. G. Bryan, F. Loayza, and E. Peláez, "A voice analysis approach for recognizing parkinson's disease patterns," *IFAC-PapersOnLine*, vol. 54, no. 15, pp. 382–387, 2021, 11th IFAC Symposium on Biological and Medical Systems BMS 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896321016918 23

[56] M. B. Er, E. Isik, and I. Isik, "Parkinson's detection based on combined cnn and lstm using enhanced speech signals with variational mode decomposition," *Biomedical Signal Processing and Control*, vol. 70, p. 103006, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809421006030 24

[57] Z. Soumaya, B. Drissi Taoufiq, N. Benayad, K. Yunus, and A. Abdelkrim, "The detection of parkinson disease using the genetic algorithm and svm classifier," *Applied Acoustics*, vol. 171, p. 107528, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X20306320 24, 42, 45

[58] N. Narendra, B. Schuller, and P. Alku, "The detection of parkinsonś disease from speech using voice source information," *IEEE/ACM Trans. Audio, Speech*

and Lang. Proc., vol. 29, p. 1925–1936, 5 2021. [Online]. Available: https://doi.org/10.1109/TASLP.2021.3078364 24, 45

[59] M. A. Mohammed, M. Elhoseny, K. H. Abdulkareem, S. A. Mostafa, and M. S. Maashi, "A multi-agent feature selection and hybrid classification model for parkinson's disease diagnosis," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 17, no. 2, 5 2021. [Online]. Available: https://doi.org/10.1145/3433180 25, 42

[60] M. Masud, P. Singh, G. S. Gaba, A. Kaur, R. Alroobaea, M. Alrashoud, and S. A. Alqahtani, "Crowd: Crow search and deep learning based feature extractor for classification of parkinson's disease," *ACM Trans. Internet Technol.*, vol. 21, no. 3, 6 2021. [Online]. Available: https://doi.org/10.1145/3418500 25

[61] H. Li, C.-M. Pun, F. Xu, L. Pan, R. Zong, H. Gao, and H. Lu, "A hybrid feature selection algorithm based on a discrete artificial bee colony for parkinson's diagnosis," *ACM Trans. Internet Technol.*, vol. 21, no. 3, 6 2021. [Online]. Available: https://doi.org/10.1145/3397161 25, 42

[62] D. V. Rao, Y. Sucharitha, D. Venkatesh, K. Mahamthy, and S. M. Yasin, "Diagnosis of parkinson's disease using principal component analysis and machine learning algorithms with vocal features," *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, pp. 200–206, 2022. 25

[63] D. Yao, W. Chi, and M. Khishe, "Parkinson's disease and cleft lip and palate of pathological speech diagnosis using deep convolutional neural networks evolved by ipwoa," *Applied Acoustics*, vol. 199, p. 109003, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X22003772 26

[64] C. Quan, K. Ren, Z. Luo, Z. Chen, and Y. Ling, "End-to-end deep learning approach for parkinson's disease detection from speech signals," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 2, pp. 556–574, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0208521622000341 26, 45

[65] M. Hireš, M. Gazda, L. Vavrek, and P. Drotár, "Voice-specific augmentations for parkinson's disease detection using deep convolutional neural network," *2022 IEEE 20th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp. 000 213–000 218, 2022. 26

[66] M. Hireš, M. Gazda, P. Drotár, N. D. Pah, M. A. Motin, and D. K. Kumar, "Convolutional neural network ensemble for parkinson's disease detection from voice recordings," *Computers in Biology and Medicine*, vol. 141, p. 105021, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482521008155 27

[67] F. Chen, C. Yang, and M. Khishe, "Diagnose parkinson's disease and cleft lip and palate using deep convolutional neural networks evolved by ip-based chimp

optimization algorithm," *Biomedical Signal Processing and Control*, vol. 77, p. 103688, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809422002105 27

[68] A. Lahoti, K. Gurugubelli, J. R. O. Arroyave, and A. K. Vuppala, "Shifted delta cepstral coefficients with rnn to improve the detection of parkinson's disease from the speech - proceedings of the 2022 fourteenth international conference on contemporary computing," p. 284–288, 2022. [Online]. Available: https://doi.org/10.1145/3549206.3549258 27

[69] G. V. D. Kumar, V. Deepa, N. Vineela, and G. Emmanuel, "Detection of parkinson's disease using lightgbm classifier," *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1292–1297, 2022. 27

[70] A. Hussain and A. Sharma, "Machine learning techniques for voice-based early detection of parkinson's disease," *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pp. 1436–1439, 2022. 28

[71] C. Laganas, D. Iakovakis, S. Hadjidimitriou, V. Charisis, S. B. Dias, S. Bostantzopoulou, Z. Katsarou, L. Klingelhoefer, H. Reichmann, D. Trivedi, K. R. Chaudhuri, and L. J. Hadjileontiadis, "Parkinson's disease detection based on running speech data from phone calls," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 5, pp. 1573–1584, 2022. 28

[72] X. Wang, X. Chen, Q. Wang, and G. Chen, "Early diagnosis of parkinson's disease with speech pronunciation features based on xgboost model," *2022 IEEE 2nd International Conference on Software Engineering and Artificial Intelligence (SEAI)*, pp. 209–213, 2022. 29, 41

[73] U. R. M, R. R. P, S. V, T. S, and R. S, "Parkinson's disease and its's stage classification using artificial intelligence," *2022 Second International Conference on Advanced Technologies in Intelligent Control, Environment, Computing Communication Engineering (ICATIECE)*, pp. 1–5, 2022. 29

[74] M. Mamun, M. I. Mahmud, M. I. Hossain, A. M. Islam, M. S. Ahammed, and M. M. Uddin, "Vocal feature guided detection of parkinson's disease using machine learning algorithms," *2022 IEEE 13th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, pp. 0566–0572, 2022. 29, 42, 45

[75] R. Islam, E. Abdel-Raheem, and M. Tarique, "Voiced features and artificial neural network to diagnose parkinson's disease patients," *2022 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, pp. 132–136, 2022. 30

[76] P. Deepa and R. Khilar, "Parkinson's disease classification from speech signal parameters using deep neural network," *2022 IEEE 4th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, pp. 44–48, 2022. 30

[77] M. Pramanik, S. Borah, P. Nandy, and R. Pradhan, "A machine leaning based parkinson's detection using acoustic features," *2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA)*, pp. 1–4, 2022. 30, 41

[78] R. Khaskhoussy and Y. B. Ayed, "Improving parkinson's disease recognition through voice analysis using deep learning," *Pattern Recognition Letters*, vol. 168, pp. 64–70, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865523000764 31

[79] K. Dabbabi, A. Kehili, and A. Cherif, "Parkinson detection using vot-mfcc combination and fully-connected deep neural network (fc- dnn) classifier," *2023 IEEE International Conference on Advanced Systems and Emergent Technologies (IC$_A SET$), pp.*$1 - -6, 2023.$ 31, 40

[80] A. M. Anter, A. W. Mohamed, M. Zhang, and Z. Zhang, "A robust intelligence regression model for monitoring parkinson's disease based on speech signals," *Future Generation Computer Systems*, vol. 147, pp. 316–327, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X23001942 32, 42

[81] G. Celik and E. Başaran, "Proposing a new approach based on convolutional neural networks and random forest for the diagnosis of parkinson's disease from speech signals," *Applied Acoustics*, vol. 211, p. 109476, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X23002748 32, 41

[82] K. A. Shastry, "An ensemble nearest neighbor boosting technique for prediction of parkinson's disease," *Healthcare Analytics*, vol. 3, p. 100181, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772442523000485 32

[83] D. Hemmerling, M. Wójcik-P□dziwiatr, P. Jaciów, B. Ziółko, and M. Igras-Cybulska, "Monitoring of parkinson's disease progression based on speech signal," *2023 6th International Conference on Information and Computer Technologies (ICICT)*, pp. 132–137, 2023. 33, 42

[84] A. Govindu and S. Palwe, "Early detection of parkinson's disease using machine learning," *Procedia Computer Science*, vol. 218, pp. 249–261, 2023, international Conference on Machine Learning and Data Engineering. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050923000078 33

[85] Y. Hao, "Speech-based detection machine learning methods on parkinson data set," *2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI)*, pp. 1362–1367, 2023. 33, 41

[86] P. Deepa and R. Khilar, "Detecting parkinson's disease from speech signals using boosting ensemble techniques," *2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)*, pp. 1–7, 2023. 34

[87] R. Guatelli, V. Aubin, M. Mora, J. Naranjo-Torres, and A. Mora-Olivari, "Detection of parkinson's disease based on spectrograms of voice recordings and extreme learning machine random weight neural networks," *Engineering Applications of Artificial Intelligence*, vol. 125, p. 106700, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197623008849 34

[88] R. Kumari and P. Ramachandran, "Cnn classification of parkinson's disease using stft spectrum of user's running speech," *2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pp. 1–4, 2023. 35, 60

[89] F. Abedinzadeh Torghabeh, S. A. Hosseini, and E. Ahmadi Moghadam, "Enhancing parkinson's disease severity assessment through voice-based wavelet scattering, optimized model selection, and weighted majority voting," *Medicine in Novel Technology and Devices*, p. 100266, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590093523000619 35, 42, 60

[90] P. Warule, S. P. Mishra, and S. Deb, "Time-frequency analysis of speech signal using wavelet synchrosqueezing transform for automatic detection of parkinson's disease," *IEEE Sensors Letters*, vol. 7, no. 10, pp. 1–4, 2023. 35

[91] M. Rahimi, Z. Al Masry, J. M. Templeton, S. Schneider, ra, and C. Poellabauer, "Beyond motor symptoms: Toward a comprehensive grading of parkinson's disease severity - proceedings of the 14th acm international conference on bioinformatics, computational biology, and health informatics," 2023. [Online]. Available: https://doi.org/10.1145/3584371.3612988 35, 42

[92] K. Bhatt, N. Jayanthi, and M. Kumar, "High-resolution superlet transform based techniques for parkinson's disease detection using speech signal," *Applied Acoustics*, vol. 214, p. 109657, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X23004553 36

[93] A. Sarlas, A. Kalafatelis, ros, Alex, G. ridis, M.-A. Kourtis, ros, and P. Trakadas, "Exploring federated learning for speech-based parkinson's disease detection - proceedings of the 18th international conference on availability, reliability and security," 2023. [Online]. Available: https://doi.org/10.1145/3600160.3605088 36

[94] T. A. Vu, N. T. T. Ha, L. M. Duc, H. Q. Huy, N. V. Dung, P. T. V. Huong, and N. T. Thanh, "A comparison of machine learning algorithms for parkinson's disease detection," *2023 12th International Conference on Control, Automation and Information Sciences (ICCAIS)*, pp. 316–321, 2023. 36

[95] N. D. R, V. A, A. R. S, A. E, and S. G, "Detecting parkinson's disease using machine learning," *2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)*, pp. 1–6, 2023. 36

[96] C. E. F. J, V. Sellam, and J. D. Rose, "Forecast the onset of parkinson's disease at all three stages using deep learning techniques," *2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI)*, pp. 1–8, 2023. 37

[97] D. Hemmerling, M. Wodzinski, J. R. Orozco-Arroyave, D. Sztaho, M. Daniol, P. Jemiolo, and M. Wojcik-Pedziwiatr, "Vision transformer for parkinson's disease classification using multilingual sustained vowel recordings," *2023 45th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pp. 1–4, 2023. 37

[98] V. Skaramagkas, A. Pentari, D. I. Fotiadis, and M. Tsiknakis, "Using the recurrence plots as indicators for the recognition of parkinson's disease through phonemes assessment," *2023 45th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pp. 1–4, 2023. 37

[99] A. Aluru, B. Amirisetty, H. Gudivada, and N. Sampath, "Parkinson's disease detection using machine learning techniques," *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–5, 2023. 37

[100] N. SrinivasaRao, D. Anusha, U. Mayuri, and S. Eswar, "An effective machine learning techniques to detect parkinson's disease," *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, pp. 564–568, 2023. 37

[101] D. Dutta, S. Pahadsingh, S. K. Routray, V. Nikitin, O. Kuziakin, and R. Saprykin, "Parkinson's disease detection using machine learning algorithms," *2023 IEEE 4th KhPI Week on Advanced Technology (KhPIWeek)*, pp. 1–6, 2023. 38

[102] J. Naanoue, R. Ayoub, F. E. Sayyadi, L. Hamawy, A. Hage-Diab, and F. Sbeity, "Parkinson's disease detection from speech analysis using deep learning," *2023 Seventh International Conference on Advances in Biomedical Engineering (ICABME)*, pp. 102–105, 2023. 38

[103] K. P. Faseela and P. Supriya, "Machine learning based parkinson's disease detection from enhanced speech," *2023 IEEE 4th Annual Flagship India Council International Subsections Conference (INDISCON)*, pp. 1–6, 2023. 38

[104] S. A. V. Vegesna, S. T. Ginnegolla, R. R. Yeruva, V. R. Arimanda, and S. Boda, "Dysphonia-based parkinson's detection using deep learning and ensemble techniques," *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1, pp. 1164–1168, 2023. 38

[105] M. K. Reddy and P. Alku, "Exemplar-based sparse representations for detection of parkinson's disease from speech," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 1386–1396, 2023. 38

[106] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. González-Rátiva, and E. Nöth, "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, Eds. Reykjavik, Iceland: European Language Resources Association (ELRA), 2014. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2014/pdf/7_Paper.pdf 39

[107] H. Jaeger, D. Trivedi, and M. Stadtschnitzer, "Mobile device voice recordings at king's college london (mdvr-kcl) from both early and advanced parkinson's disease patients and healthy controls [data set]. zenodo„" 2019. [Online]. Available: https://zenodo.org/records/2867216 39, 43, 44

[108] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [computer program]. version 6.4.02," 2023. [Online]. Available: http://www.praat.org/ 40

[109] Z. K. Abdul and A. K. Al-Talabani, "Mel frequency cepstral coefficient and its applications: A review," *IEEE Access*, vol. 10, pp. 122 136–122 158, 2022. 41

[110] D. K. D. Rohini R. mergu, "Multi-resolution speech spectrogram," *International Journal of Computer Applications*, vol. 15, no. 4, pp. 28–32, February 2011. [Online]. Available: https://ijcaonline.org/archives/volume15/number4/1937-2587/ 41

[111] [Online]. Available: https://raphaelvallat.com/antropy/build/html/index.html# 45

[112] [Online]. Available: https://disvoice.readthedocs.io/en/latest/index.html 45

[113] T. Arias-Vergara, J. C. Vásquez-Correa, and J. R. Orozco-Arroyave, "Parkinson's disease and aging: Analysis of their effect in phonation and articulation of speech," *Cognitive Computation*, vol. 9, pp. 731–748, 08 2017. 45

[114] J. Vásquez-Correa, J. Orozco-Arroyave, T. Bocklet, and E. Nöth, "Towards an automatic evaluation of the dysarthria level of patients with parkinson's disease," *Journal of Communication Disorders*, vol. 76, pp. 21–36, 11 2018. 45

[115] N. Dehak, P. Dumouchel, and P. Kenny, "Modeling prosodic features with joint factor analysis for speaker verification," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, pp. 2095–2103, 09 2007. 45

[116] E. A. Belalcázar-Bolaños, J. R. Orozco-Arroyave, J. F. Vargas-Bonilla, T. Haderlein, and E. Nöth, "Glottal flow patterns analyses for parkinson's disease detection: Acoustic and nonlinear approaches," *Lecture notes in computer science*, pp. 400–407, 2016. 45

[117] R. College, "The theory behind mann-whitney u tests kruskal-wallis anovas," www.reed.edu, (Last accessed on: 11 June 2024). [Online]. Available: https://www.reed.edu/psychology/stata/analyses/nonparametric/kruskal-wallacetheory.html 47

[118] IBM, "What is bagging?" www.ibm.com, (Accessed on : 31 May 2024). [Online]. Available: https://www.ibm.com/topics/bagging 47

[119] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, pp. 119–139, 1997. 48

[120] M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-learning-based disease diagnosis: A comprehensive review," *Healthcare*, vol. 10, p. 541, 03 2022. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8950225/ 48

Appendix A

Appendix

A.1 Features

Table 1: Antropy features

Antropy Features
pfd; se; mobility; complexity; nzc; DFA; HFD; PE; SVDE.

Table 2: Phonation features

Phonation Features
avg DFo; avg DDFo; avg Jitter; avg Shimmer; avg apq; avg ppq; avg logE; std DFo; std DDFo; std Jitter; std Shimmer; std apq; std ppq; std logE; skewness DFo; skewness DDFo; skewness Jitter; skewness Shimmer; skewness apq; skewness ppq; skewness logE; kurtosis DFo; kurtosis DDFo; kurtosis Jitter; kurtosis Shimmer; kurtosis apq; kurtosis ppq; kurtosis logE.

Table 3: Prosody features

Prosody Features
Foavg; Fostd; Fomax; Fomin; Foskew; Fokurt; Fotiltavg; Fomseavg; Fotiltstd; Fomsestd; Fotiltmax; Fomsemax; Fotiltmin; Fomsemin; Fotiltskw; Fomseskw; Fotiltku; Fomseku; 1Fomean; 1Fostd; 1Fomax; 1Fomin; 1Foskw; 1Foku; lastFoavg; lastFostd; lastFomax; lastFomin; lastFoskw; lastFoku; avgEvoiced; stdEvoiced; skwEvoiced; kurtosisEvoiced; avgtiltEvoiced; stdtiltEvoiced; skwtiltEvoiced; kurtosistiltEvoiced; avgmseEvoiced; stdmseEvoiced; skwmseEvoiced; kurtosismseEvoiced; avg1Evoiced; std1Evoiced; max1Evoiced; min1Evoiced; avglastEvoiced; stdlastEvoiced; maxlastEvoiced; minlastEvoiced; avgEunvoiced; stdEunvoiced; skwEunvoiced; kurtosisEunvoiced; avgtiltEunvoiced; stdtiltEunvoiced; skwtiltEunvoiced; kurtosistiltEunvoiced; avgmseEunvoiced; stdmseEunvoiced; skwmseEunvoiced; kurtosismseEunvoiced; avg1Eunvoiced; std1Eunvoiced; max1Eunvoiced; min1Eunvoiced; avglastEunvoiced; stdlastEunvoiced; maxlastEunvoiced; minlastEunvoiced; Vrate; avgdurvoiced; stddurvoiced; skwdurvoiced; kurtosisdurvoiced; maxdurvoiced; mindurvoiced; avgdurunvoiced; stddurunvoiced; skwdurunvoiced; kurtosisdurunvoiced; maxdurunvoiced; mindurunvoiced; avgdurpause; stddurpause; skwdurpause; kurtosisdurpause; maxdurpause; mindurpause; PVU; PU; UVU; VVU; VP; UP.

Table 4: Glottal features

Glottal Features
global avg var GCI; global avg avg NAQ; global avg std NAQ; global avg avg QOQ; global avg std QOQ; global avg avg H1H2; global avg std H1H2; global avg avg HRF; global avg std HRF; global std var GCI; global std avg NAQ; global std std NAQ; global std avg QOQ; global std std QOQ; global std avg H1H2; global std std H1H2; global std avg HRF; global std std HRF; global skewness var GCI; global skewness avg NAQ; global skewness std NAQ; global skewness avg QOQ; global skewness std QOQ; global skewness avg H1H2; global skewness std H1H2; global skewness avg HRF; global skewness std HRF; global kurtosis var GCI; global kurtosis avg NAQ; global kurtosis std NAQ; global kurtosis avg QOQ; global kurtosis std QOQ; global kurtosis avg H1H2; global kurtosis std H1H2; global kurtosis avg HRF; global kurtosis std HRF.

A.2 Mann-Whitney U-test results

Table 5: Mann-Whitney U-test result RT

Feature	p-value
mobility	0.0003987953852229031
complexity	0.0003987953852229031
SVDE	0.0003987953852229031
stddurvoiced	0.000447731504214323
se	0.0005628568009657778
UVU	0.0006302476671318381
VVU	0.0006302476671318381
PU	0.0007881081620977142
UP	0.0007881081620977142
skwdurpause	0.0009820337152557025
kurtosisdurpause	0.0009820337152557025
global avg std H1H2	0.0013014842980914616
global std avg H1H2	0.001462716586681149
nzc	0.001508745605698539
maxdurvoiced	0.0015669980457271976
avgdurvoiced	0.0016760379422127913
global avg avg H1H2	0.001841713997947157
F0tiltmax	0.0020628920220892495
std DDF0	0.0022856096219378216
stdEunvoiced	0.002530157794735169
avgmseEunvoiced	0.0027984241242595283
global skewness avg H1H2	0.0028829543049806376
std DF0	0.003092432547249798
skewness DF0	0.003092432547249798
mindurvoiced	0.004068015451793412
F0tiltstd	0.00607327961229262
avgdurpause	0.006664862485155913
F0kurt	0.0073077514061259735
global avg std QOQ	0.007430586826488202
global avg avg HRF	0.007430586826488202
Vrate	0.008005739719970317
stdtiltEvoiced	0.0087628379342051
avgtiltEvoiced	0.009583280440486 61
F0msemax	0.01047153192889743
std ppq	0.011432293437922595
F0std	0.011432293437922595
skwtiltEvoiced	0.012470507980264433
kurtosisEunvoiced	0.012470507980264433
F0max	0.01480030837184818
avgdurunvoiced	0.016103033550101876
skwdurunvoiced	0.016103033550101876
min1Evoiced	0.01901391869451818
stdmseEunvoiced	0.01901391869451818
avg logE	0.020634777744803517
global std avg NAQ	0.02108187426044241
avglastEunvoiced	0.02237481996318438
HFD	0.0242410543244147
...	...

Table 5: Mann-Whitney U-test result RT (Continued)

Feature	p-value
...	...
global std std NAQ	0.027419047263406694
Fomsemin	0.028082963269697035
global avg std HRF	0.02986864107470353
Foavg	0.033117228239539845
avgEunvoiced	0.033117228239539845
maxlastEunvoiced	0.033117228239539845
global kurtosis avg H1H2	0.03837630140502484
maxdurpause	0.03996228490949768
global skewness std H1H2	0.041635917785035836
global avg avg NAQ	0.045126948888676126
global avg std NAQ	0.045126948888676126
avg ppq	0.04799932887650669
Fotiltskw	0.04799932887650669
Fomseku	0.04799932887650669
stdtiltEunvoiced	0.04799932887650669
avg1Evoiced	0.05156862404704739
Fotiltmin	0.05535740349718898
skwmseEvoiced	0.05535740349718898
kurtosismseEvoiced	0.0593753860483302
max1Evoiced	0.0593753860483302

Table 6: Mann-Whitney U-test result SD

Feature	p-value
se	0.00013431989563453397
complexity	0.0003684598113958075
global avg avg H1H2	0.0008135586391330871
mobility	0.0008466075378945577
SVDE	0.0008466075378945577
global avg std H1H2	0.0016959408944200295
global std avg H1H2	0.0022492708379471915
global skewness std NAQ	0.0038798239601998794
global skewness avg NAQ	0.005741055261714581 5
avgEunvoiced	0.006383004613502795
global skewness avg H1H2	0.007395005936158744
F0std	0.010258505626646125
std ppq	0.011247803567915032
global kurtosis std NAQ	0.013539532326755008
global avg std HRF	0.015207411780084813
avg logE	0.017574290001300064
F0msemax	0.01916119661066136
avgtiltEvoiced	0.01916119661066136
std DF0	0.02087196226214276
global std avg HRF	0.026545094902759647
global kurtosis avg NAQ	0.026545094902759647
F0msestd	0.031566640060654905
skwtiltEvoiced	0.031566640060654905
std DDF0	0.040015062306800325
skwmseEvoiced	0.040015062306800325
global kurtosis avg H1H2	0.04029267972723356
F0max	0.04322741242115704
F0tiltmin	0.04322741242115704
DFA	0.04665501853718533
stdtiltEvoiced	0.04665501853718533
global std std HRF	0.049179484109178356
avg DF0	0.05030854143459397
F0mseavg	0.0541988643042437
global avg avg HRF	0.054206771415141776
skewness DF0	0.05833707260265279
kurtosisEunvoiced	0.05833707260265279
avglastEunvoiced	0.05833707260265279

A.3 Kruskal-Wallis H-test results

Table 7: Kruskal-Wallis H-test result RT

Feature	p-value
avgdurpause	0.0035278776415851275
PU	0.0035278776415851275
UP	0.0035278776415851275
PVU	0.005837910408457032
Vrate	0.006762764455423197
mobility	0.008659650784356255
SVDE	0.008659650784356255
kurtosisdurunvoiced	0.016282781831675537
complexity	0.02007885043010641
VP	0.021236751033704176
se	0.028894106064755672
minlastEunvoiced	0.03298294379320922
stddurpause	0.044424214426615545
stdlastEunvoiced	0.046300140000869
avgtiltEvoiced	0.046300140000869336
stdtiltEvoiced	0.051699034304572455
pfd	0.054030934819715916
stddurunvoiced	0.056002836307773855
global skewness avg QOQ	0.057720501469706226
std apq	0.06487492343775053

Table 8: Kruskal-Wallis H-test result SD

Feature	p-value
mobility	0.006633484813330927
complexity	0.006633484813330927
SVDE	0.006633484813330927
Vrate	0.023714545932908164
Fotiltmin	0.03271925073958674
HFD	0.03537794531606079
se	0.04013024458118911
global skewness avg QOQ	0.04068389045016618
Fokurt	0.06209022768063948
avgdurpause	0.06209022768063948
stdtiltEvoiced	0.0644628192795816
mindurvoiced	0.06472925035861365
PU	0.06720551273974959
UP	0.06720551273974959
global skewness std NAQ	0.07611199153198933
skwtiltEunvoiced	0.08106532239453304
1Fostd	0.08495572844504803

A.4 Box plots

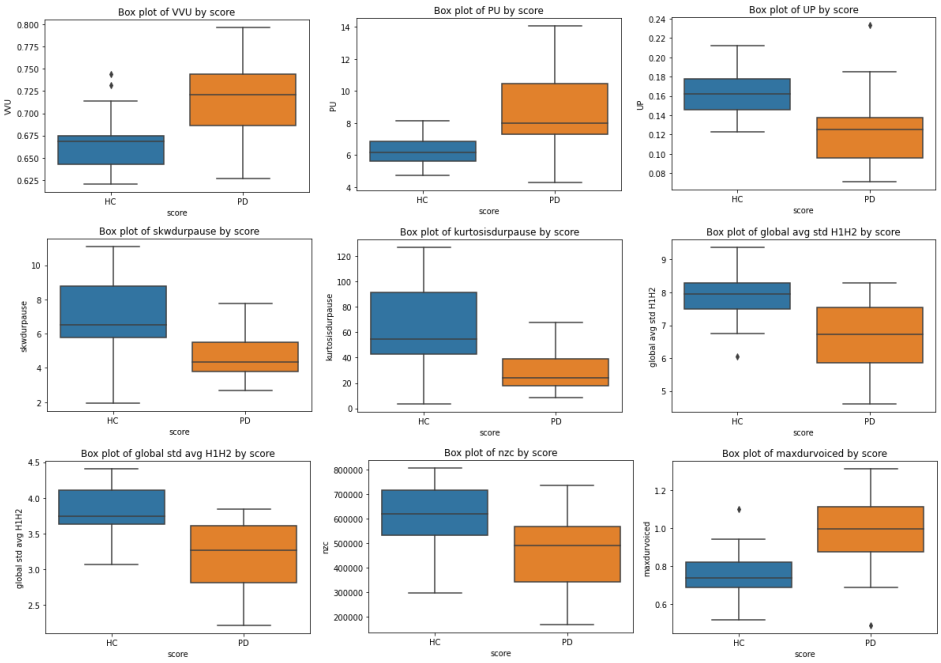

Figure 1: Box plots of features of diagnostic phase for the Read Text scenario

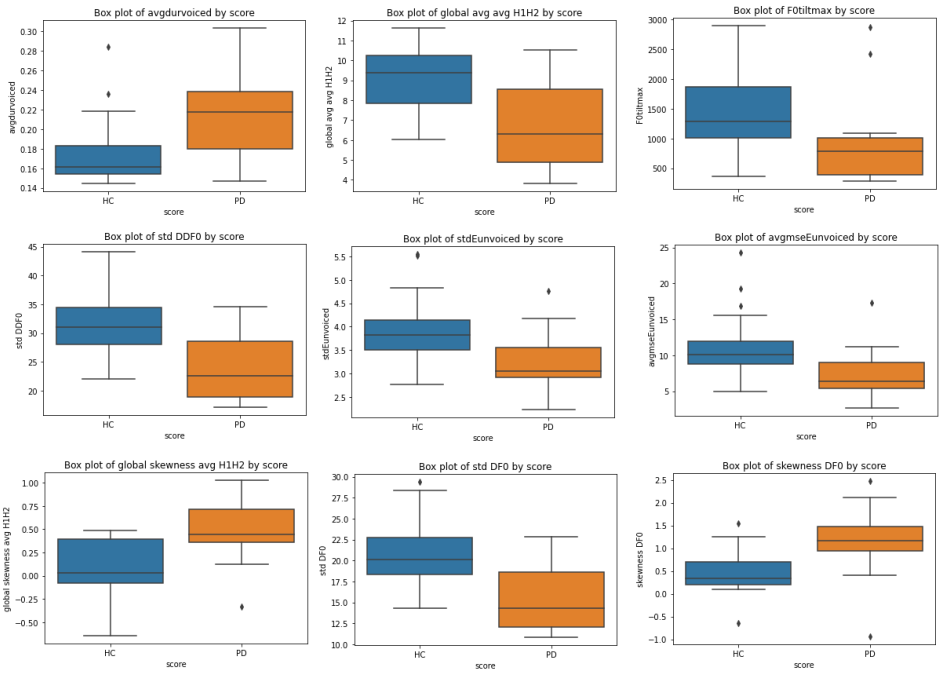

Figure 2: Box plots of features of diagnostic phase for the Read Text scenario (Continued)

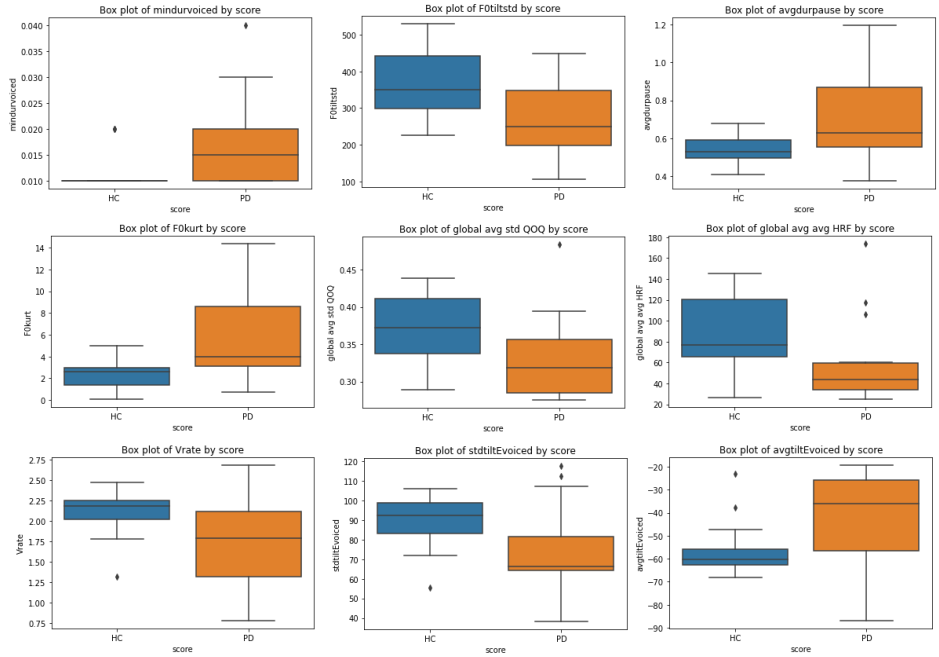

Figure 3: Box plots of features of diagnostic phase for the Read Text scenario (Continued)

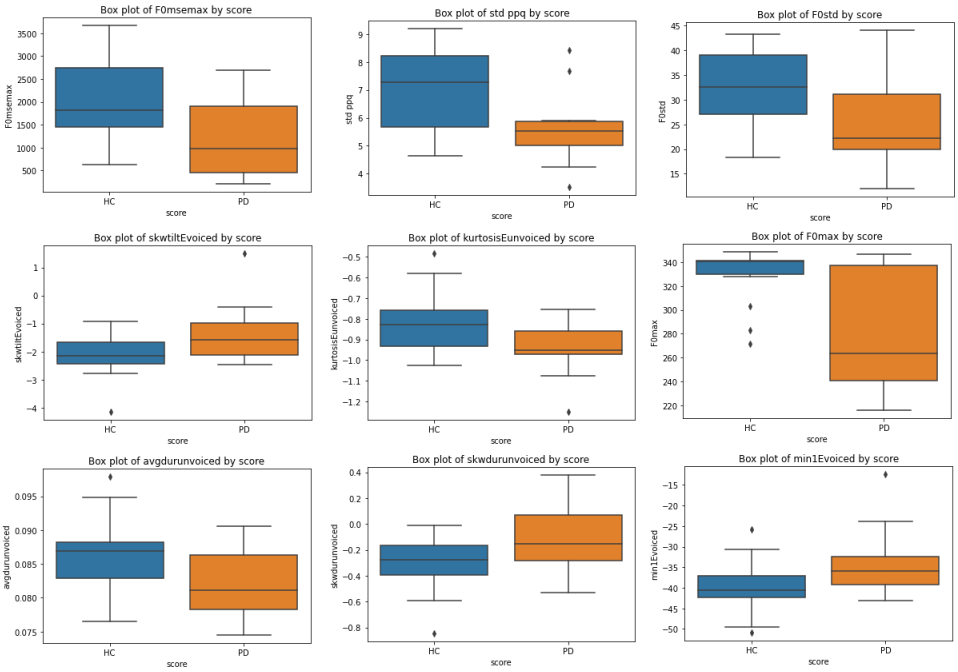

Figure 4: Box plots of features of diagnostic phase for the Read Text scenario (Continued)

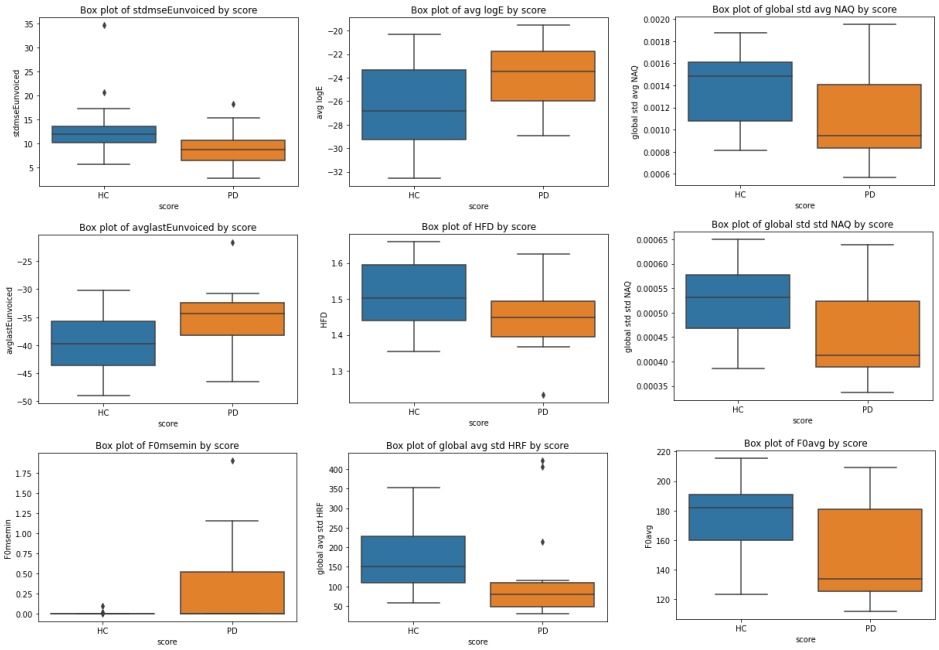

Figure 5: Box plots of features of diagnostic phase for the Read Text scenario (Continued)

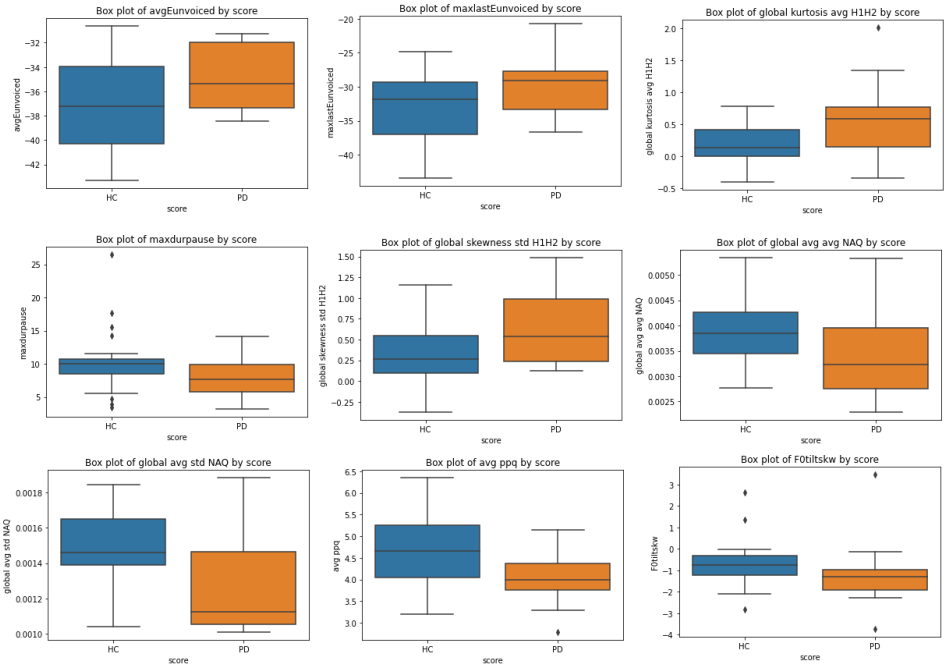

Figure 6: Box plots of features of diagnostic phase for the Read Text scenario (Continued)

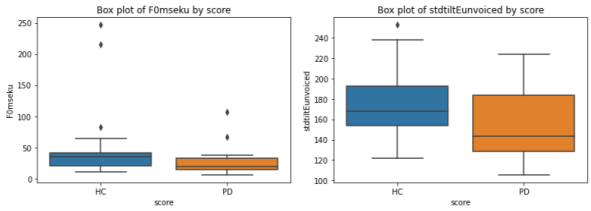

Figure 7: Box plots of features of diagnostic phase for the Read Text scenario (Continued)

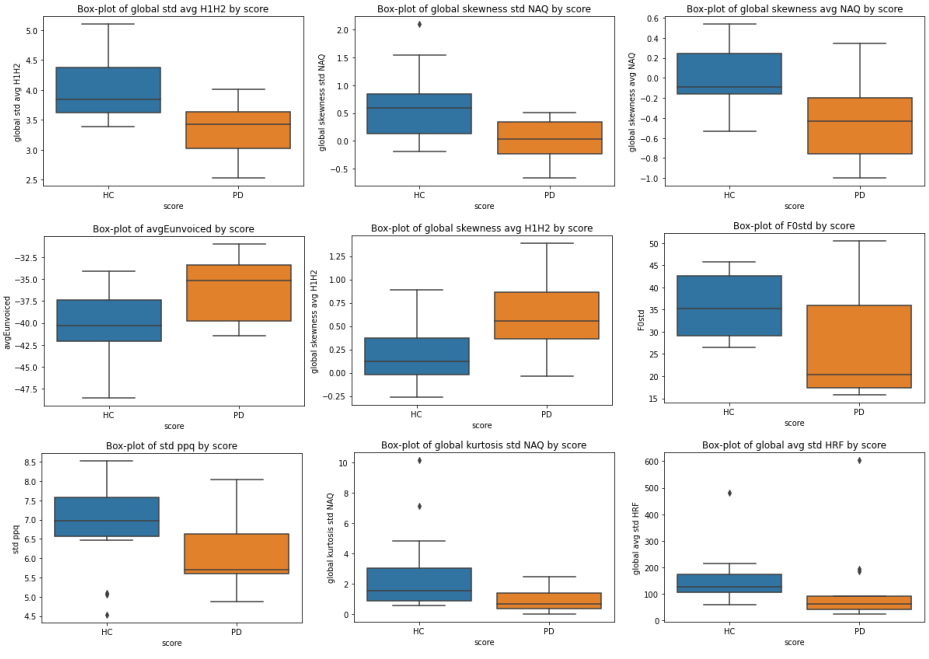

Figure 8: Box plots of features of diagnostic phase for the Spontaneous Dialogue scenario

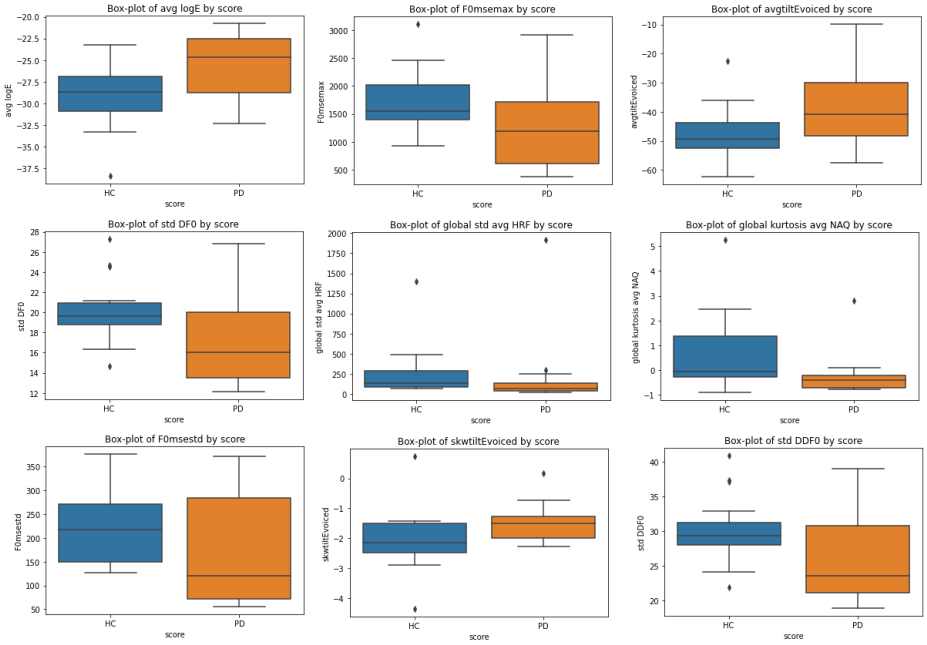

Figure 9: Box plots of features of diagnostic phase for the Spontaneous Dialogue scenario (Continued)

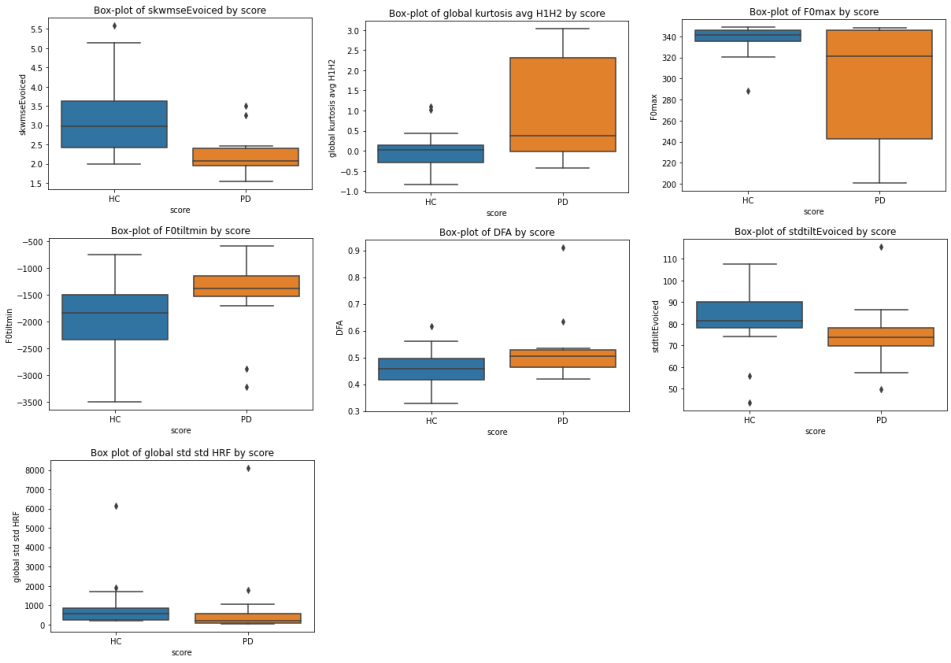

Figure 10: Box plots of features of diagnostic phase for the Spontaneous Dialogue scenario (Continued)

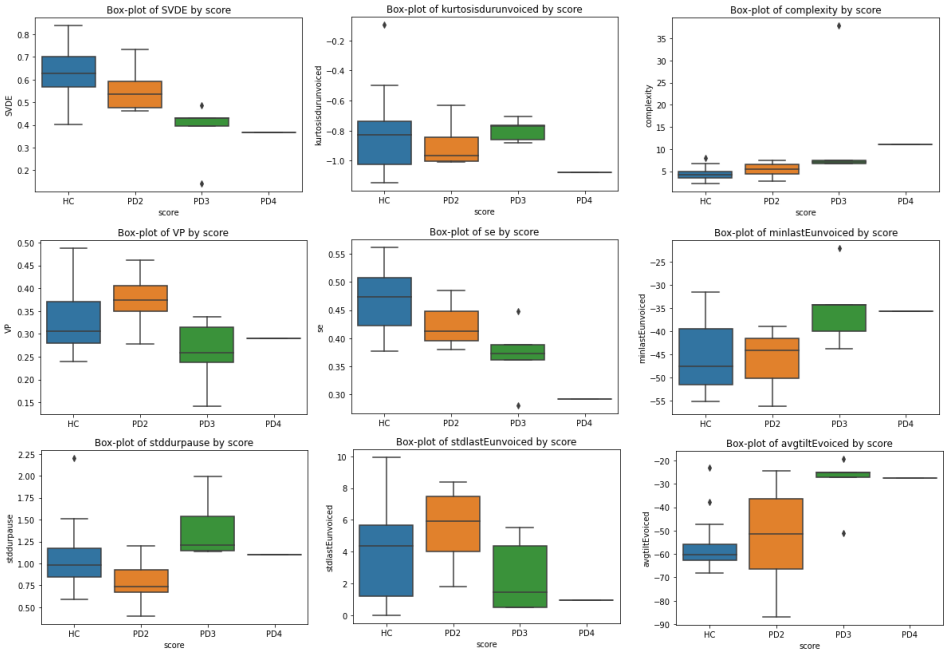

Figure 11: Box plots of features of prognostic phase for the Read Text scenario

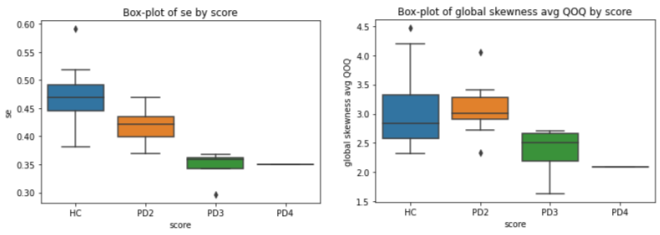

Figure 12: Box plots of features of prognostic phase for the Spontaneous Dialogue scenario

A.5 Performance results

Table 9: Performance results for diagnosis phase

	Classifier	Accuracy														
		Total	**A**	**Ph**	**Pr**	**G**	**A+Ph**	**A+Pr**	**A+G**	**Ph+Pr**	**Ph+G**	**Pr+G**	**A+Ph+Pr**	**A+Ph+G**	**A+Pr+G**	**Ph+Pr+G**
RT	**BG-KNN**	79	71	100	86	79	71	79	71	86	71	79	71	79	79	86
	BG(F)-DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	SVM-SVC	64	64	86	79	86	64	64	64	79	86	86	64	64	64	86
	AB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	BG(T)-DT	100	93	100	100	100	100	100	93	100	100	100	100	100	100	100
	RF	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	GB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	LR	93	64	100	100	100	64	100	71	100	100	100	100	79	100	100
	GNB	86	64	79	93	71	71	86	79	100	100	93	86	86	86	93
SD	**BG-KNN**	58	75	67	67	83	58	58	75	67	83	67	75	67	58	67
	BG(F)-DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	SVM-SVC	58	58	67	50	75	58	58	58	50	75	83	58	58	58	83
	AB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	BG(T)-DT	92	100	100	100	92	100	100	100	100	92	92	100	92	92	92
	RF	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	GB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	LR	100	58	83	100	92	75	100	75	100	100	100	100	92	100	100
	GNB	83	58	75	92	67	67	75	67	83	83	92	75	75	83	92

Caption: A - Antropy; Ph - Phonation; Pr - Prosody; G - Glottal; Total - A+Ph+Pr+G.

Table 9: Performance results for diagnosis phase (Continued)

	Classifier	Precision														
		Total	**A**	**Ph**	**Pr**	**G**	**A+Ph**	**A+Pr**	**A+G**	**Ph+Pr**	**Ph+G**	**Pr+G**	**A+Ph+Pr**	**A+Ph+G**	**A+Pr+G**	**Ph+Pr+G**
RT	**BG-KNN**	80/75	73/67	100	82/100	80/75	73/67	80/75	73/67	82/100	73/67	75/100	73/67	80/75	80/75	82/100
	BG(F)-DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	SVM-SVC	67/50	67/50	82/100	80/75	89/80	67/50	67/50	67/50	80/75	89/80	82/100	67/50	67/50	67/50	82/100
	AB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	BG(T)-DT	100	100/83	100	100	100	100	100	100/83	100	100	100	100	100	100	100
	RF	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	GB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	LR	100/83	64/0	100	100	100	64/0	100	73/67	100	100	100	100	80/75	100	100
	GNB	89/80	89/20	100/62	90/100	78/60	78/60	89/80	80/75	100	100	90/100	89/80	89/80	89/80	90/100
SD	**BG-KNN**	60/50	83/67	67/67	67/67	78/100	62/50	60/50	75/75	67/67	78/100	67/67	75/75	67/67	60/50	67/67
	BG(F)-DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	SVM-SVC	58/0	58/0	64/100	55/0	83/67	58/0	58/0	58/0	55/0	83/67	78/100	58/0	58/0	58/0	78/100
	AB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	BG(T)-DT	100/83	100	100	100	88/100	100	100	100	100	88/100	100/83	100	88/100	100/83	100/83
	RF	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	GB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	LR	100	58/0	78/100	100	88/100	70/100	100	83/63	100	100	100	100	88/100	100	100
	GNB	86/80	58/0	83/67	100/83	71/60	67/67	75/75	64/100	100/71	86/80	100/83	75/75	70/100	86/80	100/83

Caption: A - Antropy; Ph - Phonation; Pr - Prosody; G - Glottal; Total - A+Ph+Pr+G.
An isolated value of 100% means that this same value was obtained for all targets (HC/PD).

Table 9: Performance results for diagnosis phase (Continued)

	Classifier	Recall														
		Total	**A**	**Ph**	**Pr**	**G**	**A+Ph**	**A+Pr**	**A+G**	**Ph+Pr**	**Ph+G**	**Pr+G**	**A+Ph+Pr**	**A+Ph+G**	**A+Pr+G**	**Ph+Pr+G**
RT	**BG-KNN**	89/60	89/40	100	100/60	89/60	89/40	89/60	89/40	100/60	89/40	100/40	89/40	89/60	89/60	100/60
	BG(F)-DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	SVM-SVC	89/20	89/20	100/60	89/60	89/80	89/20	89/20	89/20	89/60	89/80	100/60	89/20	89/20	89/20	100/60
	AB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	BG(T)-DT	100	89/100	100	100	100	100	100	89/100	100	100	100	100	100	100	100
	RF	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	GB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	LR	89/100	100/0	100	100	100	100/0	100	89/40	100	100	100	100	89/60	100	100
	GNB	89/80	89/20	67/100	100/80	78/60	78/60	89/80	89/60	100	100	100/80	89/80	89/80	89/80	100/80
SD	**BG-KNN**	86/20	71/80	86/40	86/40	100/60	71/40	86/20	86/60	86/40	100/60	86/40	86/60	86/40	86/20	86/40
	BG(F)-DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	SVM-SVC	100/0	100/0	100/20	86/0	71/80	100/0	100/0	100/0	86/0	71/80	100/60	100/0	100/0	100/0	100/60
	AB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	BG(T)-DT	86/100	100	100	100	100/80	100	100	100	100	100/80	86/100	100	100/80	86/100	86/100
	RF	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	GB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	LR	100	100/0	100/60	100	100/80	100/40	100	71/80	100	100	100	100	100/80	100	100
	GNB	86/80	100/0	71/80	86/100	71/60	86/40	86/60	100/20	71/100	86/80	86/100	86/60	100/40	86/80	86/100

Caption: A - Antropy; Ph - Phonation; Pr - Prosody; G - Glottal; Total - A+Ph+Pr+G.
An isolated value of 100% means that this same value was obtained for all targets (HC/PD).

Table 10: Performance results for prognostic phase

	Classifier	Accuracy														
		Total	**A**	**Ph**	**Pr**	**G**	**A+Ph**	**A+Pr**	**A+G**	**Ph+Pr**	**Ph+G**	**Pr+G**	**A+Ph+Pr**	**A+Ph+G**	**A+Pr+G**	**Ph+Pr+G**
RT	**BG-KNN**	64	64	71	79	71	64	64	64	79	64	79	64	64	64	79
	BG(F)-DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	SVM-SVC	64	64	64	71	71	64	64	64	71	64	86	64	64	64	86
	AB	100	86	100	100	86	100	100	93	100	100	100	100	100	100	100
	BG(T)-DT	100	100	100	100	100	93	100	100	100	100	100	100	100	100	100
	RF	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	GB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	LR	100	64	100	100	100	64	100	71	100	86	100	100	79	100	100
	GNB	86	64	100	100	71	71	79	79	100	71	100	86	79	79	100
SD	**BG-KNN**	58	58	67	50	67	58	58	58	58	67	58	58	58	58	58
	BG(F)-DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	DT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	SVM-SVC	58	58	58	58	67	58	58	58	58	67	67	58	58	58	67
	AB	83	83	83	92	83	92	100	92	92	83	100	92	92	83	92
	BG(T)-DT	100	92	92	100	92	100	100	100	92	92	100	100	100	100	100
	RF	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	GB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	LR	100	58	92	100	92	67	100	83	100	100	100	100	92	100	100
	GNB	100	58	75	100	83	50	83	67	100	83	92	92	75	100	92

Caption: A - Antropy; Ph - Phonation; Pr - Prosody; G - Glottal; Total - A+Ph+Pr+G.

Table 10: Performance results for prognostic phase (Continued)

	Classifier	Precision										
		Total	**A**	**Ph**	**Pr**	**G**	**A+Ph**	**A+Pr**	**A+G**	**Ph+Pr**	**Ph+G**	**Pr+G**
RT	**BG-KNN**	73/0/50	73/0/100	75/0/100	75/100/0	69/100/0	73/0/100	73/0/50	73/0/100	75/100/0	73/0/50	75/100/0
	BG(F)-DT	100	100	100	100	100	100	100	100	100	100	100
	DT	100	100	100	100	100	100	100	100	100	100	100
	SVM-SVC	67/0/50	67/0/50	64/0/0	69/100/0	69/0/100	67/0/50	67/0/50	67/0/50	69/100/0	67/0/50	82/100/100
	AB	100	82/100/100	100	100	82/100/100	100	100	100/75/100	100	100	100
	BG(T)-DT	100	100	100	100	100	100/100/67	100	100	100	100	100
	RF	100	100	100	100	100	100	100	100	100	100	100
	GB	100	100	100	100	100	100	100	100	100	100	100
	LR	100	64/0/0	100	100	100	64/0/0	100	69/0/100	100	82/100/100	100
	GNB	89/67/100	67/0/100	100	100	88/50/50	88/40/100	80/67/100	89/50/100	100	88/40/100	100
SD	**BG-KNN**	58/0/0	64/0/0	67/100/50	55/0/0	64/100/0	64/0/0	58/0/0	64/0/0	58/0/0	64/100/0	60/50/0
	BG(F)-DT	100	100	100	100	100	100	100	100	100	100	100
	DT	100	100	100	100	100	100	100	100	100	100	100
	SVM-SVC	58/0/0	58/0/0	60/0/50	58/0/0	64/100/0	58/0/0	58/0/0	58/0/0	58/0/0	64/100/0	64/100/0
	AB	100/60/0	100/60/0	86/67/100	100/75/100	86/67/100	88/100/100	100	100/75/100	100/75/100	86/67/100	100
	BG(T)-DT	100	100/75/100	100/75/100	100	88/100/100	100	100	100	100/75/100	88/100/100	100
	RF	100	100	100	100	100	100	100	100	100	100	100
	GB	100	100	100	100	100	100	100	100	100	100	100
	LR	100	58/0/0	88/100/100	100	88/100/100	64/0/100	100	78/100/100	100	100	100
	GNB	100	58/0/0	100/60/67	100	86/67/100	57/33/50	86/100/67	83/100/40	100	86/67/100	100/75/100

Caption: A - Antropy; Ph - Phonation; Pr - Prosody; G - Glottal; Total - A+Ph+Pr+G.
An isolated value of 100% means that this same value was obtained for all targets (HC/PD2/$\geq$PD3).

Table 10: Performance results for prognostic phase (Continued)

	Classifier	Precision (Continued)			
		A+Ph+Pr	**A+Ph+G**	**A+Pr+G**	**Ph+Pr+G**
RT	**BG-KNN**	73/0/50	73/0/100	73/0/100	75/100/0
	BG(F)-DT	100	100	100	100
	DT	100	100	100	100
	SVM-SVC	67/0/50	67/0/50	67/0/50	82/100/100
	AB	100	100	100	100
	BG(T)-DT	100	100	100	100
	RF	100	100	100	100
	GB	100	100	100	100
	LR	100	75/0/100	100	100
	GNB	89/67/100	89/50/100	80/67/100	100
SD	**BG-KNN**	64/0/0	58/0/0	58/0/0	67/33/0
	BG(F)-DT	100	100	100	100
	DT	100	100	100	100
	SVM-SVC	58/0/0	58/0/0	58/0/0	64/100/0
	AB	100/75/100	100/75/100	100/60/0	88/100/100
	BG(T)-DT	100	100	100	100
	RF	100	100	100	100
	GB	100	100	100	100
	LR	100	88/100/10	100	100
	GNB	88/100/100	86/100/50	100	100/75/100

Caption: A - Antropy; Ph - Phonation; Pr - Prosody; G - Glottal; Total - A+Ph+Pr+G.
An isolated value of 100% means that this same value was obtained for all targets (HC/PD2/$\geq$PD3).

Table 10: Performance results for prognostic phase (Continued)

	Classifier	Recall										
		Total	**A**	**Ph**	**Pr**	**G**	**A+Ph**	**A+Pr**	**A+G**	**Ph+Pr**	**Ph+G**	**Pr+G**
RT	**BG-KNN**	89/0/50	89/0/50	100/0/50	100/67/0	100/33/0	89/0/50	89/0/50	89/0/50	100/67/0	89/0/50	100/67/0
	BG(F)-DT	100	100	100	100	100	100	100	100	100	100	100
	DT	100	100	100	100	100	100	100	100	100	100	100
	SVM-SVC	89/0/50	89/0/50	100/0/0	100/33/0	100/0/50	89/0/50	89/0/50	89/0/50	100/33/0	89/0/50	100/67/50
	AB	100	100/33/100	100	100	100/67/50	100	100	100/100/50	100	100	100
	BG(T)-DT	100	100	100	100	100	89/100/100	100	100	100	100	100
	RF	100	100	100	100	100	100	100	100	100	100	100
	GB	100	100	100	100	100	100	100	100	100	100	100
	LR	100	100/0/0	100	100	100	100/0/0	100	100/0/50	100	100/33/100	100
	GNB	89/67/100	89/0/50	100	100	78/67/50	78/67/50	89/67/50	89/67/50	100	78/67/50	100
SD	**BG-KNN**	100/0/0	100/0/0	86/33/50	86/0/0	100/33/0	100/0/0	100/0/0	100/0/0	100/0/0	100/33/0	86/33/0
	BG(F)-DT	100	100	100	100	100	100	100	100	100	100	100
	DT	100	100	100	100	100	100	100	100	100	100	100
	SVM-SVC	100/0/0	100/0/0	86/0/50	100/0/0	100/33/0	100/0/0	100/0/0	100/0/0	100/0/0	100/33/0	100/33/0
	AB	100/100/0	71/100/100	86/67/100	86/100/100	86/67/100	100/67/100	100	100/100/50	86/100/100	86/67/100	100
	BG(T)-DT	100	100/100/50	100/100/50	100	100/100/50	100	100	100	100/100/50	100/100/50	100
	RF	100	100	100	100	100	100	100	100	100	100	100
	GB	100	100	100	100	100	100	100	100	100	100	100
	LR	100	100/0/0	100/67/100	100	100/67/100	100/0/50	100	100/67/50	100	100	100
	GNB	100	100/0/0	57/100/100	100	86/67/100	57/33/50	86/67/100	71/33/100	100	86/67/100	86/100/100

Caption: A - Antropy; Ph - Phonation; Pr - Prosody; G - Glottal; Total - A+Ph+Pr+G.
An isolated value of 100% means that this same value was obtained for all targets (HC/PD2/≥PD3).

Table 10: Performance results for prognostic phase (Continued)

	Classifier	Recall (Continued)			
		A+Ph+Pr	**A+Ph+G**	**A+Pr+G**	**Ph+Pr+G**
RT	**BG-KNN**	89/0/50	89/0/50	89/0/50	100/67/0
	BG(F)-DT	100	100	100	100
	DT	100	100	100	100
	SVM-SVC	89/0/50	89/0/50	89/0/50	100/67/50
	AB	100	100	100	100
	BG(T)-DT	100	100	100	100
	RF	100	100	100	100
	GB	100	100	100	100
	LR	100	100/0/100	100	100
	GNB	89/67/100	89/67/50	89/67/50	100
SD	**BG-KNN**	100/0/0	100/0/0	100/0/0	86/33/0
	BG(F)-DT	100	100	100	100
	DT	100	100	100	100
	SVM-SVC	100/0/0	100/0/0	100/0/0	100/33/0
	AB	100/100/50	100/100/50	100/100/0	100/67/100
	BG(T)-DT	100	100	100	100
	RF	100	100	100	100
	GB	100	100	100	100
	LR	100	100/67/100	100	100
	GNB	100/67/100	86/33/100	100	86/100/100

Caption: A - Antropy; Ph - Phonation; Pr - Prosody; G - Glottal; Total - A+Ph+Pr+G.
An isolated value of 100% means that this same value was obtained for all targets (HC/PD2/$\geq$PD3).